

日本国特許庁  
JAPAN PATENT OFFICE

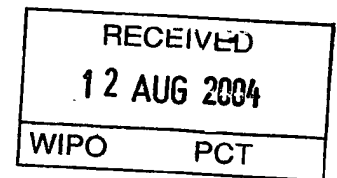
22 APR 2005  
25.11.03

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application: 2003年 9月30日

出願番号  
Application Number: 特願2003-342676  
[ST. 10/C]: [JP2003-342676]



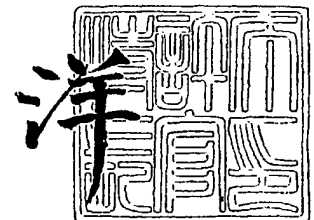
出願人  
Applicant(s): 独立行政法人産業技術総合研究所  
独立行政法人 科学技術振興機構

PRIORITY DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

2004年 7月29日

特許庁長官  
Commissioner,  
Japan Patent Office

小川



BEST AVAILABLE COPY

出証番号 出証特2003-3107785

【書類名】 特許願  
【整理番号】 K011P23  
【提出日】 平成15年 9月30日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G10L 11/00  
【発明者】  
    【住所又は居所】 茨城県つくば市東1-1-1 独立行政法人産業技術総合研究所  
                    つくばセンター内  
    【氏名】 後藤 真孝  
【特許出願人】  
    【識別番号】 301021533  
    【氏名又は名称】 独立行政法人産業技術総合研究所  
【特許出願人】  
    【識別番号】 396020800  
    【氏名又は名称】 科学技術振興事業団  
【代理人】  
    【識別番号】 100091443  
    【弁理士】  
    【氏名又は名称】 西浦 ▲嗣▼晴  
【先の出願に基づく優先権主張】  
    【出願番号】 特願2002-309519  
    【出願日】 平成14年10月24日  
【手数料の表示】  
    【予納台帳番号】 076991  
    【納付金額】 10,500円  
【その他】 国等以外のすべての者の持分の割合 050/100  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
    【包括委任状番号】 0111071

**【書類名】 特許請求の範囲****【請求項 1】**

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算ステップと、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の前記繰り返し区間を時間軸上に統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定ステップとからなることを特徴とする音楽音響データ中のサビ区間を検出する方法。

**【請求項 2】**

前記特徴量抽出ステップで求める音響特徴量は、1 オクターブの範囲に含まれる 12 の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た 12 次元クロマベクトルである請求項 1 に記載の音楽音響データ中のサビ区間を検出する方法。

**【請求項 3】**

前記類似度演算ステップでは、今回求めた前記音響特徴量と先に求めた全ての前記音響特徴量との間の前記類似度を求めることを特徴とする請求項 2 に記載の音楽音響データ中のサビ区間を検出する方法。

**【請求項 4】**

前記類似度演算ステップでは、時刻  $t$  の前記 12 次元クロマベクトルとそれよりラグ 1 ( $0 \leq 1 < t$ ) だけ過去の全ての前記 12 次元クロマベクトルとの前記類似度を求め、

前記繰り返し区間リストアップステップでは、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上前記類似度が予め定めた閾値以上ある場合には前記類似度が前記予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を前記時間軸を基準にした前記繰り返し区間としてリストアップすることを特徴とする請求項 3 に記載の音楽音響データ中のサビ区間を検出する方法。

**【請求項 5】**

統合繰り返し区間決定ステップでは、前記時間軸の共通区間に存在するリストアップした前記類似線分どうしをそれぞれグルーピングにより統合して前記統合繰り返し区間と定め、

複数の前記統合繰り返し区間を、前記共通区間の前記時間軸上の存在位置及び長さでグルーピングされる前記類似線分の前記ラグ軸で見た位置関係とに基づいて前記複数種類の前記統合繰り返し区間列に分類することを特徴とする請求項 4 に記載の音楽音響データ中のサビ区間を検出する方法。

**【請求項 6】**

統合繰り返し区間決定ステップでは、前記統合繰り返し区間に含まれない最初の繰り返し区間を補足して前記統合繰り返し区間列を作成する請求項 5 に記載の音楽音響データ中のサビ区間を検出する方法。

**【請求項 7】**

前記楽曲は転調を含んでおり、

前記特徴量抽出ステップでは、前記 12 次元クロマベクトルからなる前記音響特徴量を 1 転調幅ずつ 11 転調幅までシフトして得た転調幅の異なる 12 種類の前記音響特徴量を求め、

前記類似度演算ステップでは、今回求めた前記音響特徴量と先に求めた全ての 12 種類

の前記音響特徴量ととの間の前記類似度を、時刻  $t$  の今回の前記音響特徴量を表す前記クロマベクトルとそれよりラグ  $l$  ( $0 \leq l < t$ ) だけ過去の全ての  $12$  種類の前記音響特徴量を表す前記クロマベクトルととの間の類似度として演算し、

前記繰り返し区間リストアップステップでは、 $12$  種類の前記音響特徴量ごとに、一方の軸を時間軸  $t$  とし他方の軸をラグ  $l$  とし、予め定めた時間長さ以上前記類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を前記時間軸を基準にした前記繰り返し区間としてそれぞれ  $12$  種類のリストをリストアップすることを特徴とする請求項 1 に記載の音楽音響データ中のサビ区間を検出する方法。

【請求項 8】

統合繰り返し区間決定ステップでは、前記  $12$  種類のリストごとに、前記時間軸の共通区間に存在するリストアップした前記類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定め、

さらに前記  $12$  種類のリストについて定めた複数の前記統合繰り返し区間を、前記共通区間の前記時間軸上の存在位置及び長さ、グルーピングされる前記類似線分の前記ラグ軸で見た位置関係とに基づいて前記複数種類の転調を考慮した前記複数種類の統合繰り返し区間列に分類化することを特徴とする請求項 7 に記載の音楽音響データ中のサビ区間を検出する方法。

【請求項 9】

前記サビ区間決定ステップでは、前記統合繰り返し区間列に含まれる前記統合繰り返し区間の前記類似度の平均と、数と長さとに基づいて該統合繰り返し区間列に含まれる前記統合繰り返し区間のサビらしさを求め、最もサビらしさの高い前記統合繰り返し区間列に含まれる前記統合繰り返し区間を前記サビ区間として決定することを特徴とする請求項 1 に記載の音楽音響データ中のサビ区間を検出する方法。

【請求項 10】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して表示手段に表示する装置であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある  $1$  以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備し、

前記複数種類の統合繰り返し区間列が前記表示手段に表示され、

前記サビ区間を含む前記統合繰り返し区間列が他の前記統合繰り返し区間列とは異なる表示態様で表示されることを特徴とする音楽音響データ中のサビ区間を検出する装置。

【請求項 11】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して表示手段に表示する装置であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある  $1$  以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複

数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備することを特徴とする音楽音響データ中のサビ区間を検出する装置。

【請求項 12】

統合繰り返し区間決定手段は、前記統合繰り返し区間に含まれない最初の繰り返し区間を補足して前記統合繰り返し区間列を作成するように構成されている請求項 11 に記載の音楽音響データ中のサビ区間を検出する装置。

【請求項 13】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して前記サビ区間を再生手段により再生する装置であって

、  
前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、  
前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備し、

前記複数種類の統合繰り返し区間列が選択的に前記再生手段で再生されることを特徴とする音楽音響データ中のサビ区間を検出する装置。

【請求項 14】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法をコンピュータを用いて実現するために用いられるプログラムであって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算ステップと、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定ステップとを前記コンピュータに実行させるように構成されていることを特徴とするプログラム。

【請求項 15】

統合繰り返し区間決定ステップでは、前記統合繰り返し区間に含まれない最初の繰り返し区間を補足して前記統合繰り返し区間列を作成する請求項 14 に記載のプログラム。

## 【書類名】明細書

【発明の名称】音楽音響データ中のサビ区間を検出する方法及び装置並びに該方法を実行するためのプログラム

## 【技術分野】

## 【0001】

本発明は、市販のCD (compact disc) などに収録されている、歌曲や複数種類の楽器音を同時に含む楽曲についてのアナログまたはデジタル音楽音響信号やMIDIデータ (標準MIDIファイル) 等の各種の再生可能な音楽音響データを対象にして、サビ (chorus, refrain) 区間を検出する方法及び装置並びにこの方法をコンピュータを用いて実現するためのプログラムに関するものである。

## 【背景技術】

## 【0002】

従来のサビ検出方法の一つでは、楽曲の音響信号の代表的部分として、サビを指定した長さだけ不完全に一箇所切り出す。Logan等〔非特許文献1〕は、切り出した短いフレーム (1秒間) にその部分の特徴量に基づいてラベルを付与し、最頻出のラベルをもつフレームをサビとみなす方法を提案した。このラベルの付与には、各区間の特徴量間の類似度に基づくクラスタリングや隠れマルコフモデルを用いていた。また、Bartsch等〔非特許文献2〕は、ビートトラッキングの結果に基づいて楽曲を拍ごとの短いフレームに分割し、それらの特徴量間の類似度が、指定した一定の長さの区間に渡って最も高い箇所を、サビとして切り出す方法を提案した。また、Foote〔非特許文献3〕は、非常に短い断片 (フレーム) ごとの特徴量間の類似度に基づく境界検出の応用例として、サビが切り出せる可能性を指摘していた。

## 【0003】

一方、標準MIDIファイル等の音符相当表現を対象とした従来技術〔非特許文献4および5〕もあるが、この技術は音源分離が困難な混合音にはそのまま適用できなかった。

## 【0004】

さらに関連技術として、非特許文献6以下の公知技術がある。

【非特許文献1】Logan, B. and Chu, S.: Music Summarization Using Key Phrases, Proc. of ICASSP 2000, II-749-752 (2000).

【非特許文献2】Bartsch, M. A. and Wakefield, G. H.: To Catch A Chorus: Using Chroma-based Representations for Audio Thumbnailing, Proc. of WASPAA 2001, 15-18 (2001).

【非特許文献3】Foote, J.: Automatic Audio Segmentation Using A Measure of Audio Novelty, Proc. of ICME 2000, I-452-455 (2000).

【非特許文献4】Meek, C. and Birmingham, W. P.: The Automatic Extractor, Proc. of ISMIR 2001, 119-128 (2001).

【非特許文献5】村松純: 歌謡曲における「さび」の楽譜情報に基づく特徴抽出—小室哲哉の場合—, 情報研報 音楽情報科学, 2000-MUS-35-1, 1-6 (2000).

【非特許文献6】大津展之: 判別および最小2乗規準に基づく自動しきい値選定法, 信学論 (D), J63-D, 4, 349-356 (1980).

【非特許文献7】Shepard, R. N.: Circularity in Judgments of Relative Pitch, J. Acoust. Soc. Am., 36, 12, 2346-2353 (1964).

【非特許文献8】Wakefield, G. H.: Mathematical Representation of Joint Time-Chroma Dist

tribution, SPIE 1999, 637-645 (1999).

【非特許文献9】Savitzky, A. and Golay, M. J.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry, 36, 8, 1627-1639 (1964)

【非特許文献10】後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC研究用音楽データベース; ポピュラー音楽データベースと著作権切れ音楽データベース, 情処研報 音楽情報科学, 2001-MUS-42-6, 35-42 (2001).

【非特許文献11】van Rijsbergen, C. J.: Information Retrieval, Butterworths, second edition (1979).

【非特許文献12】平田圭二, 松田周: パピプーン: GTTMに基づく音楽要約システム, 情処研報 音楽情報科学, 2002-MUS-46-5, 29-36 (2002).

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、上記したようないずれの従来の技術では、楽曲中に何度も出現するサビのどこか一箇所だけを検出していた。また、従来の技術では、常に指定した一定の長さを切り出して提示するだけで、サビの区間がどこからどこまでかは推定していなかった。また、サビが繰り返されるときに転調することがあるが、従来の技術では、いずれも転調を考慮していない。転調後のサビ区間は、転調前のサビ区間との間の特徴量の類似度が低くなるため、サビとして検出することができなかった。

【0006】

本発明の目的は、従来の技術の問題点を克服し、楽曲中に出現するサビ区間を網羅的に検出することができる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラムを提供することにある。

【0007】

本発明の目的は、1つのサビ区間がどこからどこまでかを検出することができる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラムを提供することにある。

【0008】

本発明の他の目的は、転調されたサビ区間も検出できる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラムを提供することにある。

【0009】

本発明の他の目的は、サビ区間だけでなく、その他の繰り返し区間も表示手段に表示することができる音楽音響データ中のサビ区間を検出する装置を提供することにある。

【0010】

本発明のさらに他の目的は、サビ区間だけでなく、その他の繰り返し区間も再生することができる音楽音響データ中のサビ区間を検出する装置を提供することにある。

【課題を解決するための手段】

【0011】

サビは、楽曲全体の構造の中で、一番代表的な盛り上がる主題の部分である。通常、サビは楽曲中で最も多く繰り返され、印象に残るため、専門的な音楽の訓練を受けていない人が音楽を聴いたときでも、どこがサビであるかを容易に判断できる。さらに、サビ検出の結果は、様々な応用において有用である。例えば、多数の楽曲をブラウジングするときや、楽曲検索システムにおいて検索結果を提示するときに、サビの冒頭を短く再生（プレビュー）できると便利である（画像のサムネールの音楽版とみなせる）。また、歌声等を検索キーとした楽曲検索では、検索対象をサビ区間に限定すると精度と効率が上がる。そして本発明のサビ検出技術を実施すれば、サビ区間を自動的にインデキシングすることも

可能になる。

#### 【0012】

本発明の方法は、ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出するために、特徴量抽出ステップと、類似度演算ステップと、繰り返し区間リストアップステップと、統合繰り返し区間決定ステップと、サビ区間決定ステップとを実行する。

#### 【0013】

最初に、特徴量抽出ステップでは、音楽音響データから所定の時間単位で音響特徴量を順次求める。具体的な実施の形態では、入力されてくる音楽音響データについて、所定のサンプリング幅を持って重複しながらデータのサンプリングを行うハニング窓等のサンプリング技術を用いて、所定の時間単位（例えば80ms）で、サンプリングを行う。そしてサンプリングしたデータについて、音響特徴量を求める。音響特徴量の求め方は任意である。例えば、特徴量抽出ステップで求める音響特徴量として、1オクターブの範囲に含まれる12の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た12次元クロマベクトルを用いることができる。12次元クロマベクトルを音響特徴量として用いると、複数オクターブに渡る楽曲の特徴量を抽出できるだけでなく、転調した音楽音響データから対比が可能な特徴量として抽出することができる。

#### 【0014】

次に、類似度演算ステップでは、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める。類似度を求める際に用いる演算式は、任意であり、公知の類似度演算式のいずれを用いてもよい。そして繰り返し区間リストアップステップでは、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする。類似度演算ステップで、今回求めた音響特徴量と先に求めた全ての音響特徴量との間の類似度を求めると、リアルタイムにサビ区間を検出することが可能になる。

#### 【0015】

より具体的な、類似度演算ステップでは、時刻 $t$ のクロマベクトル（音響特徴量）とそれよりラグ $l$ （ $0 \leq l < t$ ）（ $l$ はアルファベット $l$ の小文字）だけ過去の全てのクロマベクトルとの類似度を求めることになる。この場合、繰り返し区間リストアップステップでは、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてリストアップする。なおこのリストアップは、演算上のリストアップであればよく、実際的に表示手段上にリストアップする必要はない。したがって時間軸及びラグ軸も理論上の軸であればよい。ここで「類似線分」の概念は、本願明細書において定義するものである。類似線分は、予め定めた時間長さ以上類似度が予め定めた閾値以上あるときに、閾値以上ある類似度の部分の長さに対応する時間長さを有する線分として定義される。閾値を適宜に変更または調整することにより、ノイズを除去することが可能になる。なお閾値を設けたことによりノイズは除去できるものの、本来現れるべき類似線分が現れなくなる場合もある。しかしそのような場合であっても、今回の特徴量と過去のすべての特徴量との間の類似度についての類似線分をリストアップするため、後に他の類似線分との関係から、本来現れるべき類似線分がないことを探索することができるので、リストアップの精度が下がることはない。

#### 【0016】

統合繰り返し区間決定ステップでは、リストアップされた複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定する。統合繰り返し区間決定ステップでは、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める。そして複数の統合繰り返し区間を、共通区間の長さでグルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の統合繰り返し区間に分類する。より具体的には、リストアップされた複数の繰り返し区間の相互関係は、時間軸上の共通区間に対応する過去のラグ位置に1以上の繰り返し区間（類似線分）が存在するか否かと、



そのラグ位置に対応する過去の時間帯において繰り返し区間（類似線分）が存在するか否かの関係である。これらの関係に基づいて、このステップでは、共通区間に対応する過去のラグ位置に 1 以上の繰り返し区間（類似線分）がある場合に、それらをその共通区間に繰り返し区間（類似線分）があるものと決定して、その繰り返し区間を統合繰り返し区間とする。その上で、統合繰り返し区間決定ステップでは、決定した複数の統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する。この分類化は、共通区間の長さの共通性と、共通区間に存在する繰り返し区間（類似線分）の位置関係と数との関係に基づいて行われる。この分類化により、種類の異なる繰り返し区間の構造化が実現できる。

#### 【0017】

なお統合繰り返し区間を用いると、類似度を求めた 2 番目以降の繰り返し区間に対応する統合繰り返し区間は求まるものの、最初の繰り返し区間が統合繰り返し区間列には含まれないことになる。そこで統合繰り返し区間決定ステップでは、統合繰り返し区間に含まれない最初の繰り返し区間を補足して統合繰り返し区間列を作成するようにしてもよい。

#### 【0018】

そしてサビ区間決定ステップで、複数種類の統合繰り返し区間列からサビ区間を決定する。このサビ区間決定ステップでは、例えば、統合繰り返し区間列に含まれる統合繰り返し区間の類似度の平均と、数と長さとに基づいて該統合繰り返し区間列に含まれる統合繰り返し区間のサビらしさを求める。そして、最もサビらしさの高い統合繰り返し区間列に含まれる統合繰り返し区間をサビ区間として決定する。なおサビらしさの定め方は、一つに限定されるものではなく、よりよいサビらしさの基準に基づいて判断すれば、それだけ検出精度が高まるものは勿論である。

#### 【0019】

なお楽曲が転調を含んでいる場合には、次のようにする。まず特徴量抽出ステップでは、12 次元クロマベクトルからなる音響特徴量を 1 転調幅ずつ 11 転調幅までシフトして得た転調幅の異なる 12 種類の音響特徴量を求める。次に類似度演算ステップでは、今回求めた音響特徴量と先に求めた全ての 12 種類の音響特徴量との間の類似度を、時刻  $t$  の今回の音響特徴量を表す 12 次元クロマベクトルとそれよりラグ  $l$  ( $0 \leq l < t$ ) だけ過去の全ての 12 種類の音響特徴量を表す 12 次元クロマベクトルとの間の類似度として演算する。そして繰り返し区間リストアップステップでは、12 種類の音響特徴量ごとに、一方の軸を時間軸  $t$  とし他方の軸をラグ  $l$  とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてそれぞれ 12 種類のリストをリストアップする。

#### 【0020】

統合繰り返し区間決定ステップでは、12 種類のリストごとに、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める。さらに 12 種類のリストについて定めた複数の統合繰り返し区間を共通区間の時間軸上の存在位置及び長さ、グルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の転調を考慮した複数種類の統合繰り返し区間列に分類化する。このようにすると、転調を含んだ音楽音響データであっても、転調した部分の特徴量を 11 段階の転調幅のシフトでずらして類似度を求めるため、転調した部分の特徴量を正しく抽出することができる。その結果、繰り返し区間が転調されている場合でも、同じ特徴（A メロ、B メロ、サビ）の繰り返し区間であるか否かの判定を高い精度で行うことが可能になる。

#### 【0021】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して表示手段に表示する本発明のサビ区間検出装置は、音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める類似度演算手段と、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップす

る繰り返し区間リストアップ手段と、リストアップされた複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、複数種類の統合繰り返し区間列からサビ区間を決定するサビ区間決定手段とを具備する。サビ区間を含む統合繰り返し区間列または複数種類の統合繰り返し区間列は、表示手段に表示される。そしてサビ区間を含む統合繰り返し区間列が他の統合繰り返し区間列とは異なる表示態様で表示される。このようにすると検出したサビ区間を他の繰り返し区間とは区別して明瞭に表示することができる。

#### 【0022】

なお本発明は、統合繰り返し区間列を表示手段に表示せずに、音響の再生手段でサビ区間を含む統合繰り返し区間列またはその他の統合繰り返し区間列を選択的に再生するようにしてもよいのは勿論である。

#### 【0023】

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法をコンピュータを用いて実現するために用いられるプログラムは、音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める類似度演算ステップと、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、複数種類の統合繰り返し区間列からサビ区間を決定するサビ区間決定ステップとを前記コンピュータに実行させるように構成されている。

#### 【発明の効果】

#### 【0024】

本発明によれば、楽曲中に出現するサビ区間を網羅的に検出することができる。また本発明によれば、1つのサビ区間がどこからどこまでかを検出することができる。さらに本発明によれば、転調されたサビ区間も検出できる。また本発明によれば、サビ区間だけでなく、その他の繰り返し区間も再生し且つ表示手段にそれぞれ表示することが可能である。

#### 【発明を実施するための最良の形態】

#### 【0025】

以下、本発明の実施の形態について詳細に説明する。

#### 【0026】

まず、サビ区間を検出する場合の問題点について説明する。

#### 【0027】

サビ区間の検出のためには、楽曲一曲分の音響信号データ中に含まれるすべてのサビ区間の開始点と終了点とを求める必要がある。サビは、コーラス (chorus) あるいはリフレイン (refrain) とも呼ばれる。サビは、楽曲構造上、主題 (theme) を提示している部分を指す。そしてサビは、ときには伴奏の変化やメロディーの変形を伴いながら、通常は、楽曲中で最も多く繰り返される。例えば、典型的なポピュラー音楽の楽曲構造は、

{イントロ、サビ}

( (→第1の序奏部分 (Aメロ) [→第2の序奏部分 (Bメロ)] ) × n1 → サビ ) × n2

[→間奏] [→第1の序奏部分 (Aメロ)] [→第2の序奏部分 (Bメロ)] → サビ × n3

[→間奏 → サビ × n4] [→エンディング]

のようにになっている。このようにサビは、他のメロディよりも繰り返し回数が増えて

いる。ここで、 $\{a, b\}$  は  $a$  か  $b$  かのいずれか一方、 $[a]$  は  $a$  が省略可能であることを表す記号である。そして  $n_1, n_2, n_3, n_4$  は繰り返し回数を表す正の整数である（多くの場合、 $1 \leq n_1 \leq 2, 1 \leq n_2 \leq 4, n_3 \geq 0, n_4 \geq 0$ ）。イントロ (introduction) は前奏部分、Aメロ、Bメロ (verse A, verse B) は序奏部分を指す。

#### 【0028】

楽曲中で通常、最も多く繰り返されるサビの区間を検出するには、基本的には、ある楽曲中に含まれる複数の区間の繰り返し（繰り返し区間）を見つけ出し、最も出現頻度の大きい区間をサビ区間とすればよい。しかし、「繰り返し区間」とは言っても音響信号が完全に一致する状態で区間が繰り返される場合は希である。そのため、人間にとっては容易に繰り返しと分かる場合でも、計算機にとってはその判断が難しい。その際の主要な課題は、以下のようにまとめられる。

#### 【0029】

##### 課題1：特徴量と類似度の検討

ある区間の音響信号とその区間の繰り返し区間と考えられる他の区間の音響信号とが完全に一致しない場合には、ある区間が繰り返されているということを判断するために、各区間から求めた特徴量相互間の類似度を判断しなければならない。その際、繰り返しがあると判断できるためには、繰り返す度にその区間内の音響信号の細部が多少異なっても（メロディーが変形したり、伴奏のベース、ドラム等が演奏されなくなったりしても）、各区間の特徴量間の類似度は高い必要がある。しかしながら、各区間のパワースペクトルを直接特徴量とした場合には、この類似度の判断が困難である。

#### 【0030】

##### 課題2：繰り返しの判断基準

類似度がどれくらい高ければ繰り返しとみなせるかという基準は、楽曲に依存して変わる。例えば、似た伴奏が多用される楽曲では、全体的に多くの部分の類似度が高くなる。そのため、比較する各区間の類似度がかなり高い類似度でなければ、それらの区間がサビに関連する繰り返し区間であると判断しない方がよい。逆に、サビが繰り返されるときに、伴奏が大きく変化するような楽曲では、比較する各区間の類似度がやや低くても繰り返し区間であると判断する方がよい。こうした基準を、ある楽曲に特化して人間が手作業で設定するのは容易である。しかしながら、幅広い楽曲からサビ区間を自動的に検出するためには、サビ区間の検出の基準を、現在処理中の楽曲に応じて自動的に変える必要がある。このことは、あるサビ区間の検出方法の性能を評価する場合に、その方法で数曲のサンプル曲についてサビ区間の検出ができたからといって、必ずしもその方法に汎用性があるとは限らないということを意味する。

#### 【0031】

##### 課題3：繰り返し区間の端点（開始点と終了点）の推定

サビ区間の長さ（区間長）は楽曲ごとに異なるため、各区間長と共に、どこからどこまでがサビであるかを推定しなければならない。その際、サビの前後の区間も一緒に繰り返すことがあるため、端点の推定は、楽曲中の様々な箇所の情報を統合して行う必要がある。例えば、(A B C B C C) のような構造の楽曲の場合（A, B, C はそれぞれAメロ、Bメロ、サビの区間とする）、単純に繰り返し区間を探すと、(B C) が一つのまとまった区間として見つかる。この場合、最後のCの繰り返し情報に基づいて、(B C) の内のCの区間の端点を推定する、といった処理が求められる。

#### 【0032】

##### 課題4：転調を伴う繰り返しの検出

転調後の区間は、一般に特徴量が大きく変わるために、転調前の区間との類似度が低くなり、繰り返し区間と判断するのが困難となる。特に、転調は曲の後半のサビの繰り返しで起きることが多く、そうした繰り返しを的確に判断することは、サビの検出において重要な課題である。

#### 【0033】

本発明では、以上の課題を解決しつつ、基本的に楽曲中で多く繰り返される区間をサビとして検出する。以下の実施の形態の説明においては、入力として、音楽のモノラルの音響信号を対象とし、混合音中の楽器の数や種類には特に制限を設けない。ステレオ信号の場合には、左右を混合してモノラル信号に変換するものとする。以下の実施の形態では、以下のことを仮定する。

#### 【0034】

仮定1：演奏のテンポは一定でなく変化してもよい。しかしサビの区間は、毎回ほぼ類似したテンポで、一定の長さの区間として繰り返し演奏される。その区間は長い方が望ましいが、区間長には、許容される適切な範囲（現在の実装では、7.7～40sec）がある。

#### 【0035】

仮定2：前述した楽曲構造の例の、

$(\rightarrow A \text{メロ} [\rightarrow B \text{メロ}]) \times n1 \rightarrow \text{サビ}) \times n2$

に相当するような、長い繰り返しがある場合、その末尾の部分がサビである可能性が高い（図25参照）。

#### 【0036】

仮定3：サビ区間内では、その区間の半分程度の長さの短い区間が繰り返されることが多い。そのため、ある繰り返し区間内にさらに区間の短い繰り返し区間がある場合には、その区間がサビである可能性が高い（図26参照）。

#### 【0037】

以上は、多くのポピュラー音楽に当てはまる妥当な仮定である。本実施の形態においては、上記課題と仮定を前提にしている。

#### 【0038】

図1は本発明のサビ区間検出方法で、転調を伴う楽曲中のサビ区間を検出する一実施の形態の方法の処理ステップを示すフローチャートである。

#### 【0039】

(1) 本実施の形態では、まず、音響信号（音響信号データ）を得る（ステップS1）。

。

#### 【0040】

(2) 次に、その入力音響信号の各フレームから、細部の変形の影響を受け難い12次元の特徴量（12音名各々の周波数のパワーを複数のオクターブに渡って加算した12次元クロマベクトル）を抽出する（ステップS2）。

#### 【0041】

(3) その抽出された12次元クロマベクトルの特徴量と過去の全フレームの特徴量との間の類似度を計算する（課題1に対応）（ステップS3-1）。次に、判別基準に基づく自動閾値選定法〔非特許文献6〕によって、繰り返しの判断基準を楽曲ごとに自動的に変えながら、繰り返し区間のペアをリストアップする（課題2に対応）（ステップS3-2）。そして、それらのペアを楽曲全体に渡って統合することで、繰り返し区間のグループを作り、それぞれの端点も適切に求める（課題3に対応）（ステップS3-3）。

#### 【0042】

(4) ここで、転調を考慮に入れた場合、クロマベクトルの各次元は音名に対応しているため、その転調幅に応じて次元間で値をシフトさせた転調後のクロマベクトルと、転調前のクロマベクトルとは値が近くなる。そこで、そのように12種類の転調先を考慮して、転調前後のクロマベクトルの類似度を計算する。それを出発点として、上記の繰り返し区間の検出処理も12種類分行い、それら全ての繰り返し区間を統合する（課題4に対応）（ステップS4）。

#### 【0043】

(5) 最終的に、得られた各区間のサビらしさを上記の仮定に基づいて評価する（ステップS5）。

#### 【0044】

(6) 最もサビらしい区間の一覧を出力する (ステップ S 6)。

【0045】

(7) 同時に、中間結果として得られた繰り返し構造も出力する (ステップ S 7)。

【0046】

また図 2 は、本発明のサビ区間を検出する装置の実施の形態の一例の構成の概略を示すブロック図である。この装置では、図 1 の方法も当然にして実現可能である。さらに図 3 は、図 2 の装置をコンピュータを利用して実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。図 2 の装置の構成を説明しながら、併せて図 1 のステップと図 3 のフローチャートのステップについて説明する。

【0047】

まずサンプリング手段 1 は、所定のサンプリング幅を持って重複しながらデータのサンプリングを行うハニング窓等のサンプリング技術を用いて、所定の時間単位 (例えば 80 ms) で、入力されてくる音楽音響データについてサンプリングを行う (図 3 のサンプリングステップ S T 1)。データが音響信号であれば、サンプリングされるデータは、非常に短い断片 (フレーム) の音響信号である。

【0048】

特徴量抽出手段 3 は、サンプリング手段 1 で時間単位でサンプリングしたデータについて、音響特徴量を求める (図 3 の特徴量抽出ステップ S T 2)。ここで特徴量抽出手段 3 で採用する音響特徴量の求め方は任意である。この実施の形態では、特徴量抽出ステップで求める音響特徴量として、1 オクターブの範囲に含まれる 12 の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た 12 次元クロマベクトル (chroma vector) を用いる。

【0049】

ここで図 4 及び図 5 を用いて 12 次元クロマベクトルについて説明する。クロマベクトルは、非特許文献 7 に開示されているクロマ (音名, chroma) を周波数軸として、パワーの分布を表現した特徴量である。ここでクロマベクトルは、非特許文献 8 の chroma spectrum のクロマの軸を 12 個の音名に離散化したものに近いものである。図 4 に示すように、非特許文献 7 によれば、音楽的な音高の知覚 (音楽的高さと音色的高さ) は上に昇る螺旋状の構造を持つ。そして音楽的な音高の知覚は、この螺旋を真上から見た円周上のクロマと、横から見たときの縦方向のハイト (オクターブ位置, height) の二つの次元で表現することができる。クロマベクトルでは、パワースペクトルの周波数軸がこの螺旋状の構造に沿っていると見なし、螺旋をハイト軸方向につぶして円にすることで、周波数スペクトルを円周上 (1 周が 1 オクターブ) のクロマの軸だけで表現する。つまり、異なるオクターブの同じ音名の位置のパワーを加算して、クロマ軸上のその音名の位置のパワーとする。

【0050】

本実施の形態では、図 5 に示すように、このクロマベクトルを 12 次元で表し、クロマベクトルの各次元の値が平均律の異なる音名のパワーを表すものとする。図 5 では、6 オクターブの同じ音名の位置のパワーをそれぞれ加算してクロマ軸上のその音名の位置のパワーとする状態を示している。12 次元のクロマベクトルを得るためには、まず時刻  $t$  の入力音響信号に対する短時間フーリエ変換 (STFT) を計算する。その後、短時間フーリエ変換 (STFT) で求めた演算結果を、周波数軸を対数スケールの周波数  $f$  に変換して、パワースペクトル  $\Psi_p(f, t)$  を求める。対数スケールの周波数は cent の単位で表し、Hz で表された周波数  $f_{Hz}$  を、次のように cent で表された周波数  $f_{cent}$  に変換する。

【0051】

$$f_{cent} = 1200 \log_2 [f_{Hz} / (440 \times 2^{3/12 - 5})] \quad \dots (1)$$

平均律の半音は 100 cent に相当し、1 オクターブは 1200 cent に相当する。そのため、音名  $c$  ( $c$  は  $1 \leq c \leq 12$  の整数で、クロマに対応)、オクターブ位置  $h$  (ハイトに対応) の周波数  $F_{c, h cent}$  は、

$F_{c,h} = 1200h + 100 (c-1)$   
と表せる。

### 【0052】

この対数スケール軸のパワースペクトル  $\Psi_p(f, t)$  から音名  $c$  の位置のパワーを  $O_{c,t_L}$  から  $O_{c,t_H}$  (現実の実装では、3~8) のオクターブ範囲で加算して、12次元クロマベクトルの各次元  $v_c(t)$  を下記式(3)で求める。

### 【数1】

$$v_c(t) = \sum_{h=O_{c,t_L}}^{O_{c,t_H}} \int_{-\infty}^{\infty} BPF_{c,h}(f) \Psi_p(f, t) df \quad \dots\dots\dots(3)$$

を求める。ここで、 $BPF_{c,h}(f)$  は、音名  $c$ 、オクターブ位置  $h$  の位置のパワーを通過させるバンドパスフィルタであり、下記式(4)のように、ハニング窓の形状で定義する。

### 【数2】

$$BPF_{c,h}(f) = \frac{1}{2} \left( 1 - \cos \frac{2\pi(f - (F_{c,h} - 100))}{200} \right) \quad \dots\dots\dots(4)$$

こうして得られたクロマベクトルを特徴量とすることで、繰り返す度に繰り返し区間のメロディーや伴奏が多少変わっても、繰り返し区間全体の響き(同時に鳴っている音名の構成)が類似していれば、その区間は繰り返し区間として検出できる。さらに、後述するように、類似度の工夫によって転調された繰り返し区間の検出も可能となる。

### 【0053】

なお現在作成している装置では、音響信号を標本化周波数 16 kHz、量子化ビット数 16 bit で A/D 変換している。そして窓関数  $h(t)$  として窓幅 4096 点のハニング窓を用いた短時間フーリエ変換(STFT)を、高速フーリエ変換(FFT)で計算する。高速フーリエ変換(FFT)のフレームは、1280 点ずつシフトし、すべての処理の時間単位(1フレームシフト)を 80 ms とする。

### 【0054】

図2に戻って、上記のようにして求めた特徴量は、特徴量記憶手段5に記憶される。そして類似度演算手段7は、それまでに入力された音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める(図3の類似度演算ステップST3)。類似度を求める際に用いる演算式は、任意であり、公知の類似度演算式のいずれを用いてもよい。そして繰り返し区間リストアップ手段9は、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする(図3の繰り返し区間リストアップステップST4)。

### 【0055】

類似度演算手段7では、今回求めた音響特徴量と先に求めた全ての音響特徴量との間の類似度を求めている。これによってリアルタイムにサビ区間を検出することが可能になる。具体的な類似度演算手段7では、図6及び図7に示すように、時刻  $t$  の12次元クロマベクトル(音響特徴量)とそれよりラグ  $l$  ( $0 \leq l < t$ ) ( $l$  はアルファベット  $L$  の小文字)だけ過去の全ての12次元クロマベクトルとの間の類似度を求めることになる。12次元クロマベクトル間の類似度の計算(図3のステップST3)について説明する。

### 【0056】

時刻  $t$  の12次元クロマベクトル  $v(t)$  (但しここで  $v$  はベクトル) と、それよりラグ(lag)  $l$  ( $0 \leq l \leq t$ ) だけ過去の12次元クロマベクトル  $v(t-l)$  (但しこ

ここで  $\vec{v}$  はベクトル) との類似度  $r(t, 1)$  を下記式 (5) に基づいて求める。

【数 3】

$$r(t, 1) = 1 - \frac{\left| \frac{\vec{v}(t)}{\max_e v_e(t)} - \frac{\vec{v}(t-1)}{\max_e v_e(t-1)} \right|}{\sqrt{12}} \dots\dots\dots (5)$$

上記式 (5) において、分母の  $(12)^{1/2}$  は、1 辺の長さがラグ 1 の 12 次元超立方体の対角線の長さであることを示している。上記式 (5) 中の分子中の下記式 (6) は、常にその超立方体の原点を含まない面上に位置するため、 $0 \leq r(t, 1) \leq 1$  となる。

【数 4】

$$\frac{\vec{v}(t)}{\max_e v_e(t)} \dots\dots\dots (6)$$

すなわち類似度  $r(t, 1)$  は、各時刻  $t$  のクロマベクトルを最大要素で正規化し、ラグ 1 だけ過去のクロマベクトルとユークリッド距離を計算し、その計算結果を 1 から引いた値である。

【0057】

次に、繰り返し区間リストアップ手段 9 における繰り返し区間のリストアップ (図 3 のステップ S T 4) について説明する。図 8 は、ある楽曲に対する後述する類似線分、類似度  $r(t, 1)$ 、パラメータ空間  $R a l l(t, 1)$  の概念図である。繰り返し区間リストアップ手段 9 では、図 8 に示すように、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には、類似線分を時間軸を基準にした繰り返し区間としてリストアップする。図 8 においては、類似線分を時間軸と平行に表示している。なおこのリストアップは、演算上のリストアップであればよく、実際に表示手段上にリストアップする必要はない。したがって時間軸及びラグ軸も理論上の軸であればよい。ここで「類似線分」の概念は、本願明細書において定義するものである。「類似線分」とは、予め定めた時間長さ以上類似度が予め定めた閾値以上あるときに、閾値以上ある類似度の部分の長さに対応する時間長さを有する線分として定義される。なお類似度の大きさは、類似線分に現れることはない。また閾値を適宜に変更または調整することにより、ノイズを除去することが可能になる。

【0058】

図 8 において、類似度  $r(t, 1)$  は右下半分の三角形内で定義される。実際に得られる  $r(t, 1)$  は、図 9 に示すように、ノイズを多く含み、サビに関連しない類似線分も存在して曖昧なことが多い。

【0059】

リストアップのために、類似度  $r(t, 1)$  に基づいて、どの区間が繰り返されているかを調べる。図 8 に示すように、類似度  $r(t, 1)$  を、横軸が時間軸  $t$ 、縦軸がラグ軸 1 の  $t-1$  平面に描画すると、繰り返されている区間に対応して、時間軸に平行な線分 (類似度が連続して高い領域) が現れる。そこで、時刻  $T 1$  から  $T 2$  の区間 (以下、 $[T 1, T 2]$  と表記する) に渡ってラグ軸  $L 1$  の位置に高い類似度を持つ線分を類似線分と呼び、 $[t = [T 1, T 2], l = L 1]$  で表す。これは、 $[T 1, T 2]$  と  $[T 1-L 1, T 2-L 1]$  が繰り返し区間であることを意味する。よって、 $r(t, 1)$  中の類似線分をすべて検出すれば、繰り返し区間の一覧が得られる。

【0060】

ここで類似線分の考え方について簡単に説明する。例えば、 $t-1$  平面に図 10 に示すように、繰り返し区間を示す類似線分が現れている場合を考える。図 10 の横軸の下に示

したアルファベットの表記は、それまでに入力された音響信号が A メロ → B メロ → サビ (C) → サビ (C) であることを示している。このような類似線分が現れているのは、サビ C が 2 回連続しているためである。すなわち図 11 に示すように、前のサビ C の区間と後のサビ C の区間との間の類似度は、最後のサビ C の区間と他の最初の二つの区間 (A, B) との類似度と比べて高くなるため、最後のサビ C に対応する時間位置で且つラグ 1 が前のサビ C の位置に対応する部分にサビ C と同じ時間長さの類似線分が現れるのである。さらに時間が過ぎて、図 12 のようになったと仮定する。図 12 においては、理解を容易にするために、特徴量が対比された区間を各 A, B, C のアルファベットの右下に数字で示してある。例えば「A<sub>1 2</sub>」の表示は、A 1 区間の A メロと A 2 区間の A メロの特徴量の類似度が演算されて、その類似度が高いために現れた類似線分であることを示している。同じく「C<sub>3 6</sub>」は C 3 区間のサビ区間と C 6 区間のサビ区間の特徴量の類似度が演算されて、その類似度が高いために現れた類似線分であることを示している。なお 1 つのサビ区間内においてサビの 2 度の繰り返しがあられる場合には、図 13 に示すように類似線分が現れることになる。

#### 【0061】

この線分検出をコンピュータを用いて演算により実行する場合には、画像処理においてロバストな直線検出方法として多用されるハフ (Hough) 変換を用いる。ハフ変換では、 $t-1$  平面における求めたい直線をパラメータ  $a, b$  を用いて  $l = at + b$  で表すとき、画素 ( $T, L$ ) ごとにパラメータ空間に  $b = L - aT$  の軌跡を描く (画素の輝度を累積する)。そして、多くの軌跡が交わる点 (累積値の大きい点) のパラメータを持つ直線が、画像中に存在するものとみなす。類似線分の検出の場合には、時間軸に平行な線分だけを求めればよいので上記の直線の傾きは常に 0 となり、パラメータ空間は 1 次元と単純化される。

#### 【0062】

具体的には、時刻  $t$  におけるパラメータ空間  $R_{all}(t, 1)$  は、下記式 (7) から求めることができる。

#### 【数 5】

$$R_{all}(t, 1) = \int_1^t \frac{r(\tau, 1)}{t-1} d\tau \quad \dots\dots\dots (7)$$

図 8 に示されるように、上記  $R_{all}(t, 1)$  が大きい値を持つ  $l$  の位置に類似線分が存在する可能性が高いと考える。

#### 【0063】

なお、広帯域ノイズ等に起因する各成分がほぼ等しいクロマベクトルからは、他のクロマベクトルへの距離が比較的近くなってしまう傾向があり、 $r(t, 1)$  中に類似度の高い直線 (以下、ノイズ直線と呼ぶ) として現れることがある。このノイズ直線は、 $t-1$  平面において、時間軸に垂直 (上下) 方向、あるいは、斜め右上・左下方向に現れる。そこで、前処理として式 (7) の計算前にノイズ直線の抑制を行う。まず、各  $r(t, 1)$  において、右、左、上、下、右上、左下の 6 方向の近傍区間の平均値を計算し、その最大値と最小値を求める。そして、右か左の方向の近傍区間の平均値が最大のときは、類似線分の一部とみなして、強調するために  $r(t, 1)$  から最小値を引く。その他の方向の近傍空間の平均値が最大のときは、ノイズ直線の一部とみなして、抑制するために  $r(t, 1)$  から最大値を引く。このようにして求めた  $R_{all}(t, 1)$  は、図 14 の右側に示すような線図となる。

#### 【0064】

上記のように、 $R_{all}(t, 1)$  を求めた後の類似線分の検出は、以下の手順 1 及び 2 に従って行う。



## 【0065】

手順1: 線分候補ピークの検出

図14の右側の線図に示される  $R_{all}(t, l)$  中の十分に高いピークを、線分候補ピークとして検出する。まず、 $R_{all}(t, l)$  の  $lag$  軸方向のピークを、2次多項式適合による平滑化微分を用いたピーク検出〔非特許文献9〕により求める。具体的には、下記式(8)で求める  $R_{all}(t, l)$  の平滑化微分が正から負に変わる箇所をピークとする ( $KSize = 0.32sec$ )。

【数6】

$$\sum_{w=-KSize}^{KSize} w R_{all}(t, l+w) \dots\dots\dots (8)$$

ただし、このピーク検出の前に、 $R_{all}(t, l)$  の  $lag$  軸方向に、2階のカーディナルB-スプライン関数を重み関数とする移動平均によってスムージングをかけたものを引いて、 $r(t, l)$  のノイズ成分等の蓄積による大局的な変動を取り除いておく〔 $R_{all}(t, l)$  にハイパスフィルタをかけることに相当する〕。

## 【0066】

次に、こうして得られたピークの集合から、ある閾値より大きいピークのみを、線分候補ピークとして選ぶ。前述の課題2で述べたように、この閾値は楽曲ごとに適切な値が異なるため、楽曲に基づいて自動的に変える必要がある。そこで、 $R_{all}(t, l)$  のピーク値を閾値によって二つのクラスに分けるときに、クラス分離度を最大とする判別基準に基づく自動閾値選定法〔非特許文献6〕を用いる。この自動閾値選定法は、図15に示すように閾値によって二つのクラスに分けるという考え方を採用している。ここでは、クラス分離度としてクラス間分散

$$\sigma^2_B = \omega_1 \omega_2 (\mu_1 - \mu_2)^2 \dots\dots\dots (9)$$

を最大とする閾値を求める。ただし、 $\omega_1 \omega_2$  は、閾値によって分けられた二つのクラスの生起確率 (各クラスのピーク個数/全体のピーク個数)、 $\mu_1$ 、 $\mu_2$  は、各クラスのピーク値の平均である。

## 【0067】

手順2: 類似線分の探索

図16に示すように、各線分候補ピークの  $lag$  軸上の位置  $l$  において、類似度  $r(t, l)$  の時間軸方向を一次関数とみなして、それが連続して十分高い区間を探索し、類似線分とする。

## 【0068】

まず、 $r(t, l)$  の時間軸方向に、2階のカーディナルB-スプライン関数を重み関数とする移動平均によってスムージングをかけた  $r_{smooth}(t, l)$  を求める。次に、 $r_{smooth}(t, l)$  中で、ある閾値を連続して越えているすべての区間のうち、一定の長さ (6.4sec) 以上のものを類似線分として求める。この閾値も、上記の判別基準に基づく自動閾値選定法により定める。ただし、今度はピーク値を扱うのではなく、ピーク値が高い上位5個の線分候補ピークを選び、それらのラグ  $l$  の位置の  $r_{smooth}(\tau, l)$  ( $1 \leq \tau \leq t$ ) がとる値を二つのクラスに分ける。

## 【0069】

上記のようにしてリストアップされた繰り返し区間のリストは、図2に示すリスト記憶手段11に記憶される。統合繰り返し区間決定手段13は、リスト記憶手段11に記憶されたリストから複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定する。そして統合繰り返し区間決定手段13は、さらに決定した複数の統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する。

## 【0070】

この統合繰返し区間決定ステップ（図3のST5）では、図17に示すように、前述のt-1平面における時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰返し区間RPと定める。そして複数の統合繰返し区間RPを、共通区間の位置及び長さでグルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の統合繰返し区間列に分類する。

## 【0071】

より具体的には、図17に示すように、リストアップされた複数の繰返し区間 $C_{12} \sim C_{56}$ （類似線分）の相互関係は、時間軸上の共通区間に対応する過去のラグ位置に1以上の繰返し区間 $C_{12} \sim C_{56}$ （類似線分）が存在するか否かと、そのラグ位置に対応する過去の時間帯において繰返し区間（類似線分）が存在するか否かの関係である。例えば、C6の共通区間に繰返し区間を示す類似線分 $C_{16}$ がある場合、その繰返し区間のラグ位置に対応する過去のラグ位置にも類似線分 $C_{12}$ があるという関係である。これらの関係に基づいて、このステップでは、共通区間に対応する過去のラグ位置に1以上の繰返し区間（類似線分）がある場合に、それらをグルーピング化してその共通区間に繰返し区間（類似線分）があるものと決定し、その繰返し区間を統合繰返し区間RP2, RP5, RP6等とする。ただし、図18に示すように、本来存在している最初の繰返し区間に対応しては、過去の時間帯には類似線分は無い。そのため最初の繰返し区間に対応する統合繰返し区間RP1については、最初の統合繰返し区間RP2とその共通区間に存在する類似線分 $C_{12}$ を基準にして補足する。なおこの補足は、プログラミングによって簡単に実現できる。このようにして1種類の統合繰返し区間列が作られる。

## 【0072】

図19は、共通区間の長さが長い場合の統合繰返し区間RP1及びRP2の列を作る場合の状況を示している。図20は、図13のようにサビ区間に2回の繰返しがあるために、統合繰返し区間RPの共通区間の長さが図17及び図18の統合繰返し区間列を構成する統合繰返し区間の1/2になる場合の状況を示している。このようにして統合繰返し区間決定ステップでは、決定した複数の統合繰返し区間を複数種類の統合繰返し区間列に分類化する。この分類化は、共通区間の長さの共通性と、共通区間に存在する繰返し区間（類似線分）の位置関係と数との関係に基づいて行われる。

## 【0073】

統合繰返し区間決定手段13により決定した、統合繰返し区間は統合繰返し区間列として統合繰返し区間記憶手段15に記憶される。図21は、統合繰返し区間列を表示手段18に表示した一例を示している。

## 【0074】

前述の統合繰返し区間決定手段13で実行されている統合処理をコンピュータを用いてより高い精度で実行する場合のより具体的な手順について説明する。前述の各類似線分は、ある区間が二回繰返されていることだけを表すため、例えばAとA'のペア、A'とA''のペアが、それぞれ繰返し区間として検出されたときには、それらを一つの繰返し区間のグループとして統合する必要がある。ここで、ある区間がn回（ $n \geq 3$ ）繰返されている場合には、もれなく検出されるとすると、 $n(n-1)/2$ 本の類似線分が検出される。そこで、同じ区間の繰返しを表す類似線分をグルーピングし、繰返し区間を統合する。さらに、もれていた類似線分の検出や、得られた類似線分が適切であるかの検証も行う。

## 【0075】

この統合処理は、以下の手順で実現する。

## 【0076】

手順1：類似線分のグルーピング

ほぼ同じ区間の類似線分を、一つのグループにまとめる。各グループ $\phi_i = [ [Ts_i, Te_i], Ti ]$ は、区間 $[Ts_i, Te_i]$ と、類似線分（区間が決まれば、線分候補

補ピークと対応する)のlag値 $v_{ij}$ の集合 $\Upsilon_i = \{v_{ij} \mid j=1, 2, \dots, M_i\}$  ( $M_i$ はピークの個数)で表される。そして、この類似線分のグループ $\phi_i$ の集合を、 $\Phi = \{\phi_i \mid i=1, 2, \dots, N\}$  ( $N$ はグループの個数)とする。

【0077】

手順2: 線分候補ピークの再検出

グループ $\phi_i$ ごとに、区間 $[T_{si}, T_{ei}]$ 内の類似度 $r(t, l)$ に基づいて、類似線分を改めて求めなおす。これにより、もれていた類似線分の検出ができ、例えば、図8で、ABCCの繰り返しに相当する長い類似線分上で、Cの繰り返しに相当する類似線分2か所が得られていなくても、この処理で検出されることが期待できる。

【0078】

まず、 $[T_{si}, T_{ei}]$ 内に限定して、ハフ変換のパラメータ空間 $R[T_{si}, T_{ei}](l)$  ( $0 \leq l < T_{si}$ )を下記式(10)で作成する。

【数7】

$$R[T_{si}, T_{ei}](l) = \int_{T_{si}}^{T_{ei}} \frac{r(r, l)}{T_{ei} - T_{si}} dr \quad \dots\dots\dots(10)$$

次に、前述の線分候補ピークの検出と同様に、平滑化微分を用いたピーク検出を行い( $K_{size} = 2.8_{sec}$ )、自動閾値選定法で定めた閾値を越えた線分候補ピークのlag値 $v_{ij}$ の集合を、改めて $\Upsilon_i$ とする。

【0079】

自動閾値選定法では、 $\Phi$ の全グループの区間における $R[T_{si}, T_{ei}](l)$ のピーク値を、二つのクラスに分けるようにする。

【0080】

手順3: 類似線分の適切さの検証1

サビと無関係な類似線分からなるグループ $\phi_i$ 、あるいは、 $\Upsilon_i$ の中で無関係な線分と考えられるピークを削除する。

【0081】

似た伴奏の繰り返しが多用される楽曲の場合サビと関係ない線分候補ピークが $R[T_{si}, T_{ei}](l)$ に等間隔に多く現れる傾向がある。

【0082】

そこで、 $R[T_{si}, T_{ei}](l)$ に対して平滑化微分を用いたピーク検出を行い、一定間隔(間隔は任意)で連続して並ぶ高いピークの個数が10個より多いときサビと無関係な類似線分からなるグループだと判断し、そのグループを $\Phi$ から削除する。

【0083】

また、一定間隔で連続して並ぶ低いピークの個数が5個より多いとき、サビと無関係な線分候補ピークだと判断し、その一連のピークを $\Upsilon_i$ から削除する。

【0084】

手順4: 類似線分の適切さの検証2

$\Upsilon_i$ の中には、区間 $[T_{si}, T_{ei}]$ の一部分だけ類似度が高いピークが含まれることがあるため、そうした類似度の変動の大きいピークを削除する。そこで、当該区間の $r_{smooth}(\tau, l)$ の標準偏差を求め、ある閾値より大きいものは $\Upsilon_i$ から削除する。この閾値は、 $\phi_i$ の中で、上記で求めた類似線分に対応する線分候補ピークは信頼できると考え、それらのピークでの上記標準偏差の最大値を定数倍(1.4倍)して定める。

【0085】

手順5: 類似線分の間隔の考慮

繰り返し区間が重ならないようにするために、lag軸上で隣接する類似線分(線分候補ピーク)の間隔を、線分の長さ $T_{ei} - T_{si}$ 以上とする必要がある。そこで、線分の長さより狭い間隔を持つ二つのピークのいずれかを、全体として高いピーク集合が残るよ

うに削除し、すべての間隔が類似線分の長さ以上になるようにする。

#### 【0086】

手順6: 共通区間を持つグループを統合

$T_i$  の各ピークについて、その lag 値  $v_{ij}$  だけの過去の区間  $[T_{si-v_{ij}}, T_{ei-v_{ij}}]$  のグループがあるかを探索し、発見したら統合する。統合処理では、発見したグループのすべてのピークを、対応する lag 値の場所に持つように、 $T_i$  に線分候補ピークを追加する。発見したグループ自体は削除する。

#### 【0087】

さらに、区間  $[T_{si-v_{ij}}, T_{ei-v_{ij}}]$  に一致する線分候補ピークを持つグループ  $T_k$  (グループの区間自体は異なる) があるかも探索し、発見したら統合するか判断する。この場合、 $T_k$  の過半数のピークが  $T_i$  に含まれていれば、上記同様の統合処理を行う。含まれていなければ、 $T_i$  と  $T_k$  で同じ区間を指しているピークを比較し、低い方を削除する。上記で実際に統合がなされたら、後処理として手順5の処理を再び行う。

#### 【0088】

次に、転調を伴う繰り返しの検出 (図1のステップS4) について説明する。以上述べてきた処理は転調を考慮していなかった。しかし上記の処理は、以下のように転調を扱える処理へと容易に拡張できる。図22に示すように、転調前と転調後の12次元クロマベクトルは異なる。そこで特徴量抽出ステップ (図1のステップS2) では、図23に示すように、12次元クロマベクトルからなる音響特徴量を1転調幅ずつ11転調幅までシフトして得た転調幅の異なる12種類の音響特徴量を求める。次に類似度演算ステップ (図1のステップS3-1) では、今回求めた音響特徴量と先に求めた全ての12種類の音響特徴量との間の類似度を、時刻  $t$  の今回の音響特徴量を表す12次元クロマベクトルとそれよりラグ  $l$  ( $0 \leq l < t$ ) だけ過去の全ての12種類の音響特徴量を表す12次元クロマベクトルとの間の類似度として演算する。そして繰り返し区間リストアップステップ (図1のステップS3-2) では、図24に示すように、12種類の音響特徴量ごとに、一方の軸を時間軸  $t$  とし他方の軸をラグ  $l$  とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてそれぞれ12種類のリストをリストアップする。統合繰り返し区間決定ステップ (図1のステップS3-3及びS4) では、12種類のリストごとに、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める (S3-3)。さらに12種類のリストについて定めた複数の統合繰り返し区間を共通区間の時間軸上の存在位置及び長さ、グルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の転調を考慮した複数種類の統合繰り返し区間列に分類化する (S4)。このようにすると、転調を含んだ音楽音響データであっても、転調した部分の特徴量を1段階の転調幅のシフトでずらして類似度を求めるため、転調した部分の特徴量を正しく抽出することができる。

#### 【0089】

楽曲が転調を含んでいる場合に、これをコンピュータを用いてより具体的に処理する場合には、上記の処理を以下のとおりにする。ここで、転調は平均律の半音  $t_r$  個分上の調へ変わることによって表すことにする。 $t_r$  は  $0, 1, \dots, 11$  の12種類の値を取るものとする。 $t_r = 0$  は転調しないことを意味し、 $t_r = 10$  は半音10個分上か、全音分下へ転調することを意味する。

#### 【0090】

12次元クロマベクトル  $v(t)$  (ここで  $v$  はベクトル) は、各次元  $v_c(t)$  の値を次元間で  $t_r$  個分だけシフトさせることで、転調を表現できる特長を持つ。具体的には、ある演奏の12次元クロマベクトルを  $v(t)$  (ここで  $v$  はベクトル) とし、それを  $t_r$  個上へ転調した演奏の12次元クロマベクトルを  $v(t)'$  (ここで  $v$  はベクトル) とすると、

$$v(t) \cong S^{t_r} v(t)' \quad \dots (11)$$

となる。

【0091】

ただし、Sはシフト行列で、以下の式(12)のように12次正方行列を一つ右にシフトした行列として定義される。

【数8】

$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \quad \text{.....(12)}$$

転調を伴う繰り返しの検出の処理手順を以下に述べる。まず、クロマベクトルのこの特長を利用し、 $t_r$ ごとの12種類の類似度 $r_{t_r}(t, 1)$ を下記式(13)と定義しな

【数9】

$$r_{t_r}(t, 1) = 1 - \frac{\left| \frac{S^T \vec{v}(t)}{\max_c v_c(t)} - \frac{\vec{v}(t-1)}{\max_c v_c(t-1)} \right|}{\sqrt{12}} \quad \text{.....(13)}$$

次に、それぞれの類似度 $r_{t_r}(t, 1)$ に対して、前述した繰り返し区間のリストアップをする。ただし、自動閾値選定法は $t_r=0$ のときだけ適用し、他の $t_r$ では、 $t_r=0$ で定めた閾値を用いる。これにより、転調のない曲で、 $t_r=0$ 以外のときに類似線分が誤検出されにくくなる。そして、こうして得られた各 $t_r$ ごとの類似度と類似線分に対して、前述の統合処理を行う。その結果、 $t_r$ ごとに別々の類似線分のグループ $\phi_{t_r}$ 、 $i$ の集合 $\Phi_{t_r}$ が得られる。そこで前述した、共通区間を持つグループの統合の処理を、 $t_r$ 間にまたがって行う（異なる $t_r$ に対して共通区間を持つグループを探索する）ことで、転調を含む繰り返し区間を一つのグループとして統合する。ただし、前出の処理では「 $r_k$ の過半数のピークが $r_i$ に含まれていれば、上記同様の統合処理を行う」とあるが、ここでは常に統合処理を行う。

【0092】

以下、異なる $t_r$ から得られたグループも合わせて、 $\Phi = \{\phi_i\}$ で表す。転調区間が後から分かるように、どの $t_r$ から統合されたかという情報は保存しておく。

【0093】

図2に戻って、サビ区間決定手段17では、統合繰り返し区間記憶手段15に記憶された統合繰り返し区間列からサビ区間を決定する。なお図2の例では、サビ区間を含む統合繰り返し区間列または複数種類の統合繰り返し区間列は、表示手段18に表示される（図27参照）。そしてサビ区間を含む統合繰り返し区間列が他の統合繰り返し区間列とは異なる表示態様で表示される。このようにすると検出したサビ区間を他の繰り返し区間とは区別して明瞭に表示することができる。なおこの例では、統合繰り返し区間列を、表示手段18に表示させながら選択手段21で選択して、音響の再生手段23でサビ区間を含む統合繰り返し区間列またはその他の統合繰り返し区間列を選択的に再生することができる。

。

【0094】

図1及び図3のサビ区間決定ステップ(S5、ST6)では、例えば、統合繰り返し区間列に含まれる統合繰り返し区間の類似度の平均と、統合繰り返し区間の数と長さに基づいて統合繰り返し区間列に含まれる統合繰り返し区間のサビらしさを求める。そして、最もサビらしさの高い統合繰り返し区間列に含まれる統合繰り返し区間をサビ区間として選択する。最初に図25及び26を用いて説明した前述の仮定1乃至仮定3を満たす統合繰り返し区間は、一般的にはサビらしさが高い。

#### 【0095】

上記の仮定を考慮して、コンピュータを用いてサビ区間を自動的に選択する方法について以下に説明する。前述の類似線分のグループの集合Φの中から、ある一つのグループをサビ区間として選ぶ。そのために、各グループ $\phi_i$ のサビらしさ $v_i$ を、類似線分の平均類似度や上記した仮定に基づいて評価し、最もサビらしさ $v_i$ の高いグループをサビ区間であると判定する。その準備として、グループごとに、類似線分(線分候補ピーク $v_{ij}$ )をそれが指す二つの区間へ展開し、すべての繰り返し区間 $[P_{sij}, P_{eij}]$ とその信頼度 $\lambda_{ij}$ のペアの集合を下記式(14)により求める。

#### 【0096】

$$\Lambda_i = \{ [P_{sij}, P_{eij}], \lambda_{ij} \mid j=1, 2, \dots, M_i+1 \} \quad \dots (14)$$

ここで、 $[P_{sij}, P_{eij}] = [T_{si} - v_{ij}, T_{ei} - v_{ij}]$ とし、信頼度 $\lambda_{ij}$ は、対応する類似線分における類似度 $r_{t,r}(t, l)$ の平均とする。ただし、 $j = M_i + 1$ のときは、下記式(15)のようになる。

#### 【数10】

$$[P_{sij}, P_{eij}] = [T_{si}, T_{ei}], \lambda_{ij} = \max_{k=1}^{M_i} \lambda_{ik} \quad \dots (15)$$

サビらしさ $v_i$ は、以下の手順で評価する。

#### 【0097】

(1) 仮定2を満たす統合繰り返し区間の信頼度を増加

仮定2で述べたAメロ～サビに相当するような十分に長い統合繰り返し区間(50sec以上)を持つグループ(統合繰り返し区間列) $\phi_h$ に関して、その各区間の終了点 $P_{ehk}$ とほぼ等しい終了点 $P_{eij}$ を持つ区間が他のグループ(他の統合繰り返し区間列)にあるか探索する。発見されれば、発見されたその統合繰り返し区間がサビである可能性が高いと考え、その信頼度 $\lambda_{ij}$ を2倍する。

#### 【0098】

(2) 仮定3を満たす統合繰り返し区間の信頼度を増加

サビとして適切な区間長の範囲(仮定1)の統合繰り返し区間 $[P_{sij}, P_{eij}]$ に関して、その区間の半分程度の短い統合繰り返し区間が前半と後半に一つずつ存在するか調べる。存在する場合には、それら二つの区間の信頼度の平均の半分を、元の区間の信頼度 $\lambda_{ij}$ に加える。

#### 【0099】

(3) サビらしさを算出

上記で得られた信頼度に基づき、サビらしさを下記式(16)で算出する。

#### 【数11】

$$v_i = \left( \sum_{j=1}^{M_i+1} \lambda_{ij} \right) \log \frac{T_{ei} - T_{si}}{D_{i..n}} \quad \dots (16)$$

上記式(16)において、 $\Sigma$ の項は、グループ(統合繰り返し区間列) $\phi_i$ 中にある統

合繰り返し区間の数が多いほど、また、それらの信頼度が高いほど、サビらしさが高いことを意味する。 $\log$ の項は、そのグループ（統合繰り返し区間列）に含まれる統合繰り返し区間が長いほど、サビらしさが高いことを意味する。定数  $D_{len}$  は予備実験の結果から 1.4 sec とした。

#### 【0100】

最終的に、サビとして適切な区間長の範囲（仮定1）を持つグループの中で、下記式（17）によって決まる集合  $\Delta_m$  中の区間  $[P_{smj}, P_{emj}]$  を、サビ区間とする。

#### 【数12】

$$m = \underset{i}{\operatorname{argmax}} v_i \quad \dots\dots\dots (17)$$

ここで後処理として、隣接する  $P_{smj}$  の最小間隔を求め、区間長が最小間隔となるように  $P_{emj}$  を移動して各区間を広げ、隙間を埋める。これは、本来はサビ区間が連続して隙間がないにも関わらず、得られた繰り返し区間では隙間が空いてしまうことがあるからである。ただし、埋める隙間が大きすぎるとき（1.2 sec 以上で区間長の半分より広いとき）は埋めない。

#### 【0101】

図3に示すように、上記のようにサビ区間を決定したら（ステップST6）、その結果を図2の表示手段18にリアルタイムで表示する（ステップST7）。そして、音楽音響データの全データについて上記の処理が終了するまで、上記処理が繰り返される（ステップST8）。

#### 【0102】

次に、上記実施の形態のサビ区間検出装置の実際とこの装置を用いた実験結果について説明する。実験では、音楽音響信号を音楽音響データとして入力した。そして検出したサビ区間の一覧をリアルタイムに出力することとした。装置は、刻一刻と、過去の音響信号中でサビ区間と考えられる区間の一覧（リスト）を求め、中間結果として得られた繰り返し構造（繰り返し区間の一覧  $\Delta_i$ ）と共に出力し続ける。この出力を視覚化した例を図27に示す。図27において、横軸は時間軸（sec）で楽曲全体を表示しており、上半分がパワー変化、下半分の最上段がサビ区間を含む統合繰り返し区間列の一览（最後のサビは転調を伴う）、下5段が他の統合繰り返し区間列の繰り返し構造を表す。

#### 【0103】

評価実験として、「RWC研究用音楽データベース：ポピュラー音楽」〔非特許文献10〕の100曲（RWC-MDB-P-2001, No. 1~100）を対象に、本装置のサビ検出性能を調べた。1曲すべてを入力し終わった時点で、サビ区間として検出されたものを対象に評価する。この正誤を判定するためには、基準となる正解のサビ区間を人間が手作業で指定する必要がある。そこで、楽曲を分割して各部にサビ、Aメロ、Bメロ、間奏等をラベリングできる、楽曲構造ラベリング用エディタを開発した。ラベリングでは、相対的な調の移動幅（曲の先頭の調に対して半音何個分上か）も正解に付与する。

#### 【0104】

こうして作成した正解に基づき、各曲に対する出力結果の区間と正解のサビ区間がどれくらい重なっているかを、再現率（recall rate）、適合率（precision rate）、および両者を統合したF値（F-measure）〔非特許文献11〕の観点から評価した。以下に定義を示す。

#### 【0105】

再現率（R）＝正しく検出したサビ区間の長さの合計／正解のサビ区間の長さの合計

適合率（P）＝正しく検出したサビ区間の長さの合計／検出した区間の長さの合計

F値＝ $(\beta^2 + 1) PR / (\beta^2 P + R)$  （ $\beta = 1$  を使用）

ただし、転調を伴う場合には、相対的な調の移動幅が正解と一致したときだけ、正しく

検出したと判断したとして、F 値が 0.75 以上のとき、その曲のサビ区間を正しく得られた（正答した）と判定した。

【0106】

評価結果として、100 曲中の正答曲数を表 1 に示す。

【表 1】

	条件（使用：○，未使用：×）			
転調区間の検出 仮定 2, 3 の使用	○ ○	× ○	○ ×	× ×
正答曲数	80 曲	74 曲	72 曲	68 曲

本装置の性能は一番左の 80 曲（80 曲の平均 F 値は 0.938）である。誤検出は、サビの繰り返しが他の箇所の繰り返しより多くなかったり、曲中ほとんどが類似伴奏の繰り返しだったりしたのが主な原因だった。100 曲中には、サビに転調のある曲が 10 曲含まれているが、そのうち 9 曲は検出できていた。前述の転調を伴う繰り返しの検出をやめた場合、左から二番目のように性能が落ちた。一方、仮定 2、3 に基づく信頼度の増加をやめた場合は、右二つのようにさらに性能が落ちた。サビの繰り返しで伴奏やメロディーに大幅な変化を伴う曲は 22 曲あったが、そのうち 21 曲は検出できており、その中で変化を伴うサビ自体は 16 曲で検出できていた。

【0107】

本発明は、基本的に楽曲中で最も多く繰り返される区間をサビとして検出する。その際、様々な区間の繰り返しを楽曲全体の情報を統合しながら調べることで、従来実現されていなかった、すべてのサビ区間の開始点・終了点の一覧を得ることを可能にした。また、転調後でも繰り返しと判断できるような、クロマベクトル間の類似度を導入したことで、サビの転調も検出できるようになった。RWC 研究用音楽データベース（RWC-MDB-P-2001）100 曲を用いて評価した結果、80 曲正答でき、実世界の音響信号中のサビ区間が検出できることが確認された。

【0108】

なお、本発明は音楽要約〔非特許文献 12〕とも関連しており、本発明の装置を楽曲の要約結果としてサビ区間を提示する音楽要約方法と捉えることもできる。さらに、サビ区間よりも長い区間の要約が必要なきには、中間結果として得られた繰り返し構造を用いることで、楽曲全体の冗長性を減らした要約の提示も可能となる。例えば、中間結果として（A メロ→B メロ→サビ）の繰り返しが捉えられているときは、それを提示できる。

【0109】

この実験では、ポピュラー音楽を用いて評価したが、本発明は他の音楽ジャンルにも適用できる可能性を持つ。実際に、数曲のクラシック音楽に適用したところ、その楽曲で最も代表的な主題が提示される部分を求めることができた。

【0110】

なお、本発明は上記実施例に限定されるものではなく、本発明の趣旨に基づいて種々の変形が可能であり、これらを本発明の範囲から排除するものではない。例えば、音響特徴量として、クロマベクトル以外に、周波数スペクトル、MFCC（Mel-Frequency Cepstrum Coefficients）等を用いてもよい。それらの微分値もさらに音響特徴量として加えることも可能である。また、音響特徴量間の類似度として、以下の三つ等も考えられる。



【数 1 3】

$$r(t, l) = 1 - \frac{\left| \frac{\vec{v}(t)}{\sum_c v_c(t)} - \frac{\vec{v}(t-l)}{\sum_c v_c(t-l)} \right|}{\sqrt{2}}$$

【数 1 4】

$$r(t, l) = 1 - \frac{\left| \frac{\vec{v}(t)}{|\vec{v}(t)|} - \frac{\vec{v}(t-l)}{|\vec{v}(t-l)|} \right|}{\sqrt{2}}$$

【数 1 5】

$$r(t, l) = \frac{\vec{v}(t) \cdot \vec{v}(t-l)}{|\vec{v}(t)| |\vec{v}(t-l)|}$$

さらに、本発明は入力を音響信号でなくMIDI信号とする場合にも適用でき、その場合には、音響特徴量の代わりにMIDI信号もしくはMIDI信号特徴量を用い、類似度としてはそれらのMIDI信号もしくはMIDI信号特徴量間の距離に基づく類似度を用いればよい。

【産業上の利用可能性】

【0111】

以上、詳細に説明したように、本発明によれば、音楽CD(compact disc)等による実世界の複雑な混合音からサビ区間を検出し、各サビの区間の開始点と終了点の一覧を求めることができるだけでなく、転調を伴うサビ区間を検出することも可能である。その際、楽曲全体の中に含まれる様々な繰り返し構造(複数の統合繰り返し区間列)に基づいてサビ区間を検出する。さらに、楽曲全体の中に含まれる様々な繰り返し構造に基づいてサビを検出するため、その中間結果として、繰り返し構造の一覧も同時に得ることができる。

【図面の簡単な説明】

【0112】

【図1】本発明のサビ区間検出方法で、転調を伴う楽曲中のサビ区間を検出する一実施の形態の方法の処理ステップを示すフローチャートである。

【図2】本発明のサビ区間を検出する装置の実施の形態の一例の構成の概略を示すブロック図である。

【図3】図2の装置をコンピュータを利用して実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。

【図4】螺旋状の音高知覚を説明するための図である。

【図5】12次元クロマベクトルを説明するために用いる図である。

【図6】類似度の演算の考え方を説明するために用いる図である。

【図7】類似度の演算の考え方を説明するために用いる図である。

【図8】ある楽曲に対する類似線分、類似度 $r(t, l)$ 、パラメータ空間 $Rall(t, l)$ の概念図である。

【図9】実際に得られる類似線分の一例を示す図である。

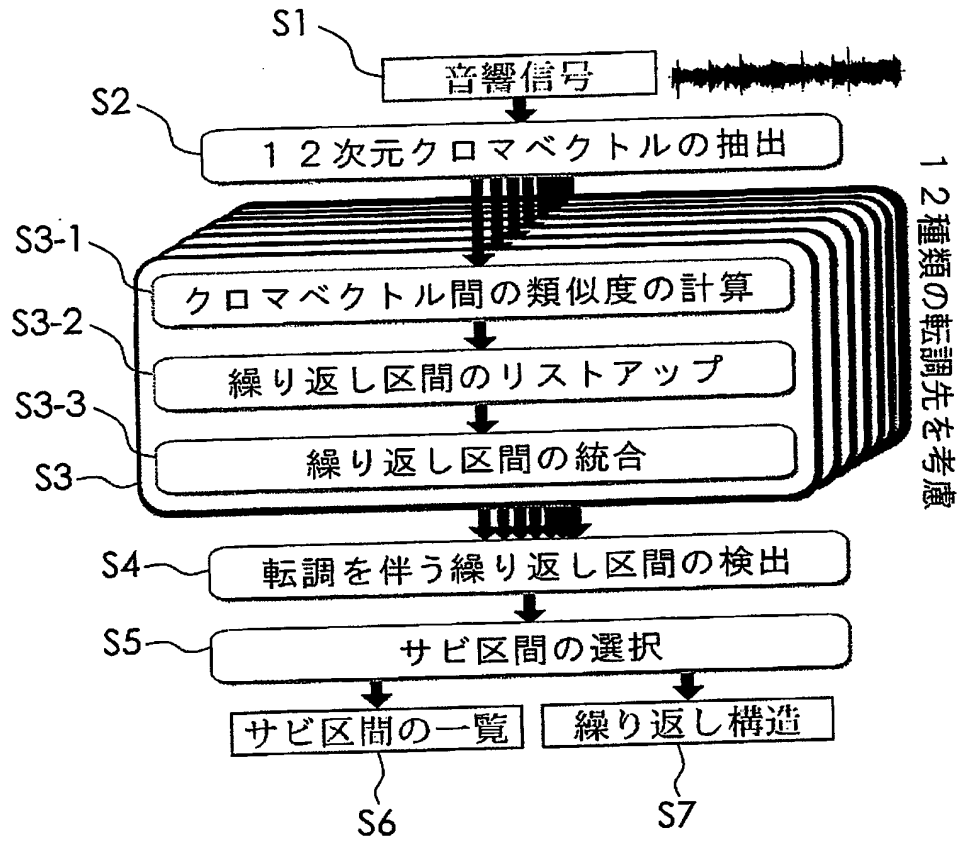
- 【図 10】類似線分の考え方を説明するために用いる図である。
- 【図 11】類似線分の考え方を説明するために用いる図である。
- 【図 12】類似線分の考え方を説明するために用いる図である。
- 【図 13】類似線分の考え方を説明するために用いる図である。
- 【図 14】類似線分を求める際の閾値の定め方を説明するために用いる図である。
- 【図 15】類似線分を求める際の閾値の定め方を説明するために用いる図である。
- 【図 16】類似線分の抽出方法を説明するために用いる図である。
- 【図 17】繰り返し区間の統合化を説明するために用いる図である。
- 【図 18】繰り返し区間の統合化を説明するために用いる図である。
- 【図 19】繰り返し区間の統合化の例を示す図である。
- 【図 20】繰り返し区間の統合化の例を示す図である。
- 【図 21】統合繰り返し区間列の表示例を示す図である。
- 【図 22】あるサビの転調前後での 12 次元クロマベクトルの違いを示す図である。
- 【図 23】転調に対処するためのシフト処理を説明するために用いる図である。
- 【図 24】転調処理のために 12 種類のリストを作成することを示す図である。
- 【図 25】サビ区間の選定の仮定の一例を説明するために用いる図である。
- 【図 26】サビ区間の選定の仮定の一例を説明するために用いる図である。
- 【図 27】RWC-MDB-P-2001, No. 18 の楽曲終了時点での正しいサビ検出結果を示す図である。

【符号の説明】

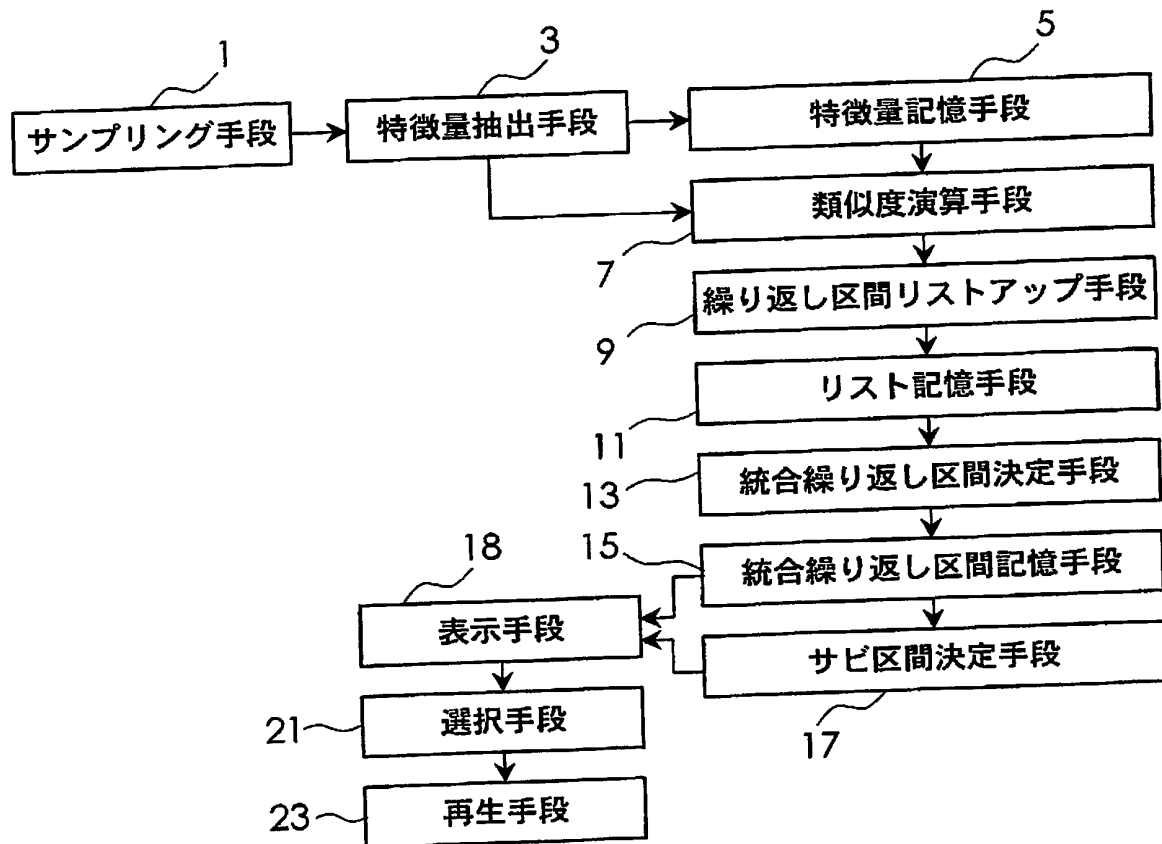
【0113】

- 1 サンプリング手段
- 3 特徴量抽出手段
- 5 特徴量記憶手段
- 7 類似度演算手段
- 9 繰り返し区間リストアップ手段
- 11 リスト記憶手段
- 13 統合繰り返し区間決定手段
- 15 統合繰り返し区間記憶手段
- 17 サビ区間決定手段
- 18 表示手段
- 21 選択手段
- 23 再生手段

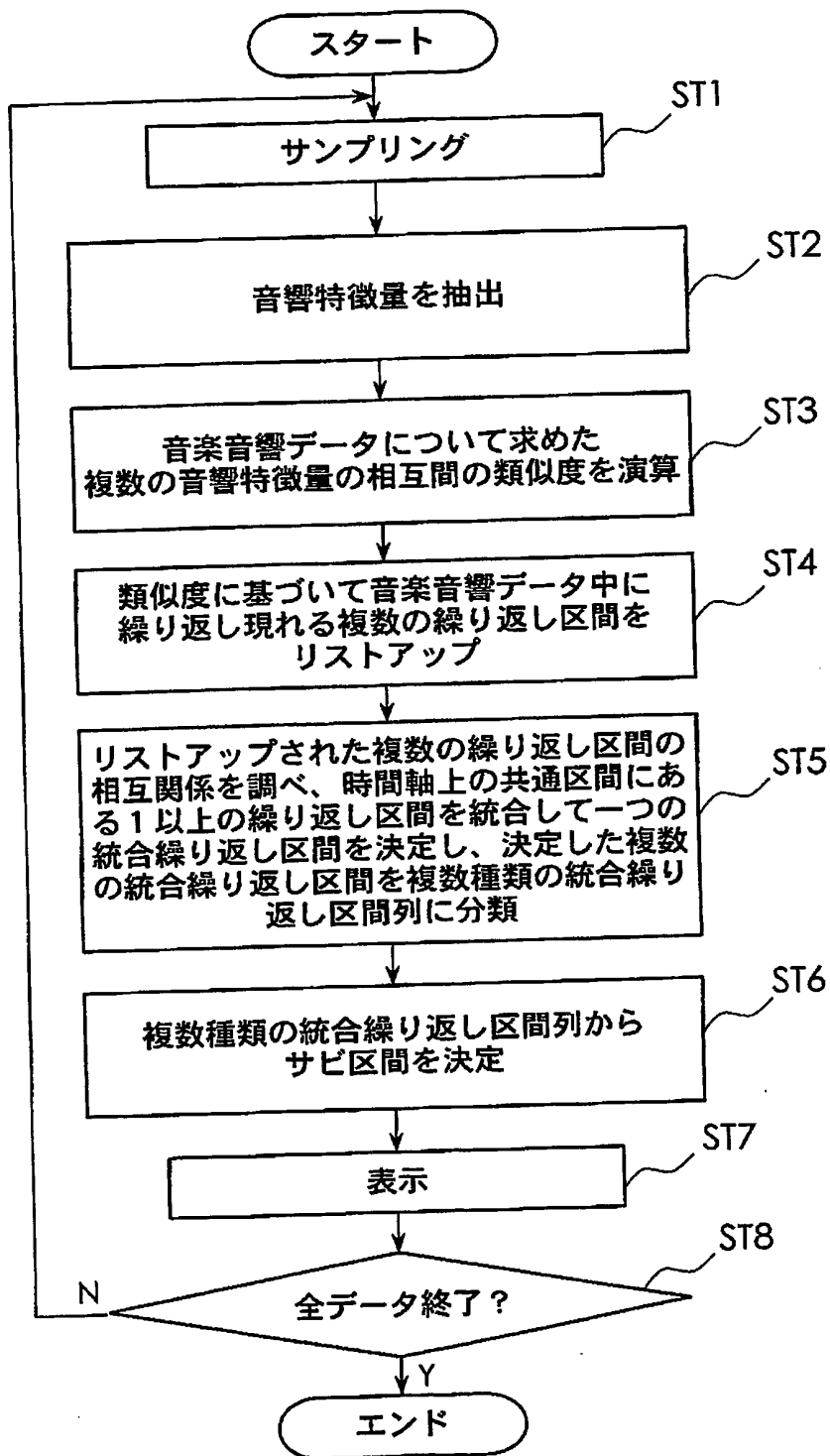
【書類名】 図面  
【図1】



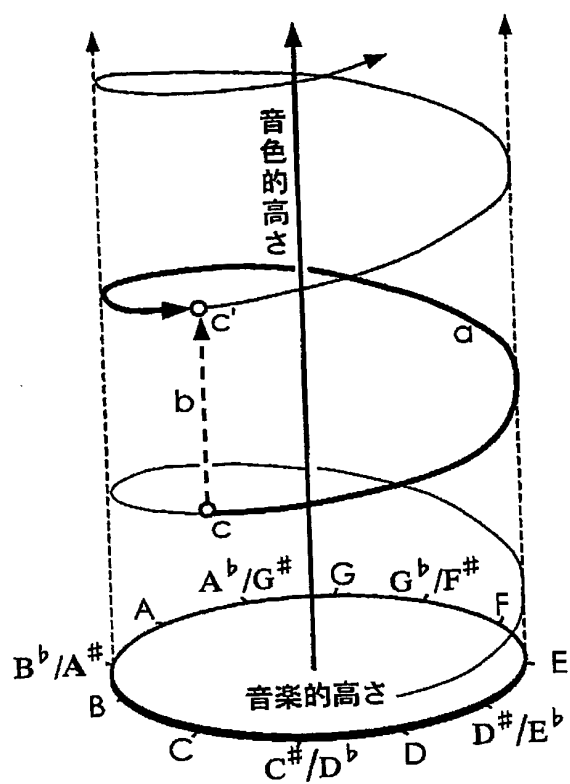
【図 2】



【図3】

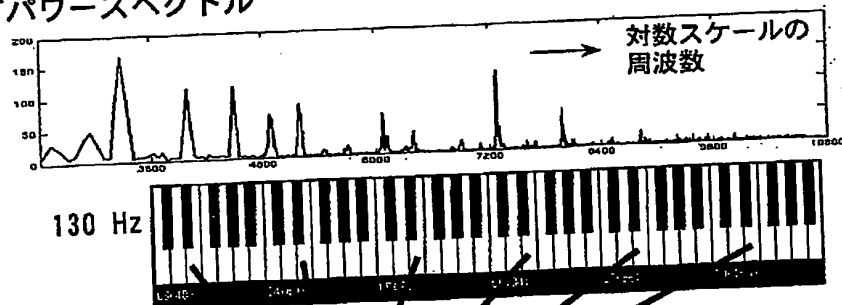


【図 4】

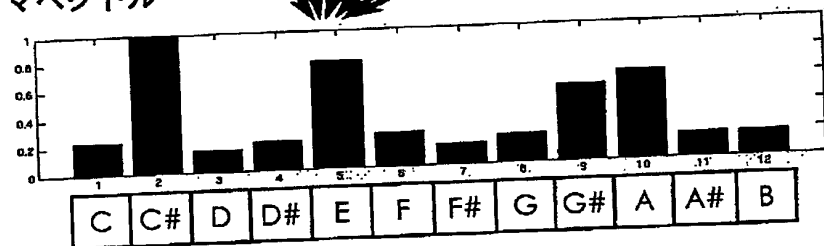


【図 5】

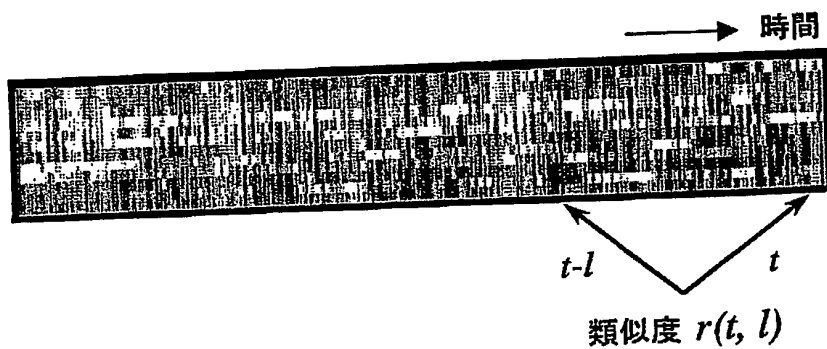
STFTパワースペクトル



クロマベクトル



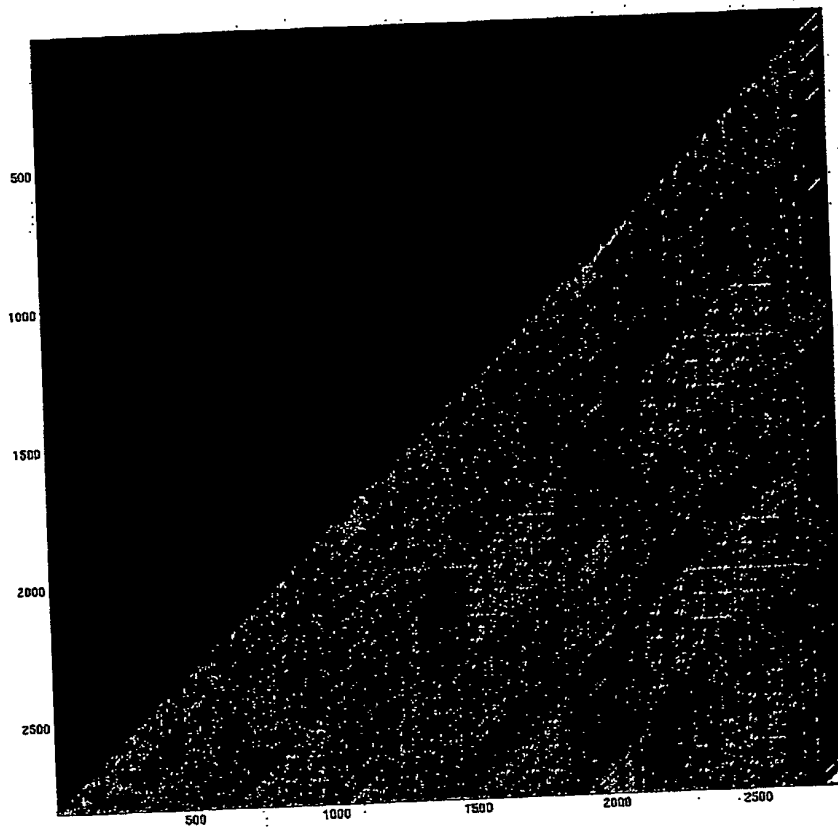
【図 6】



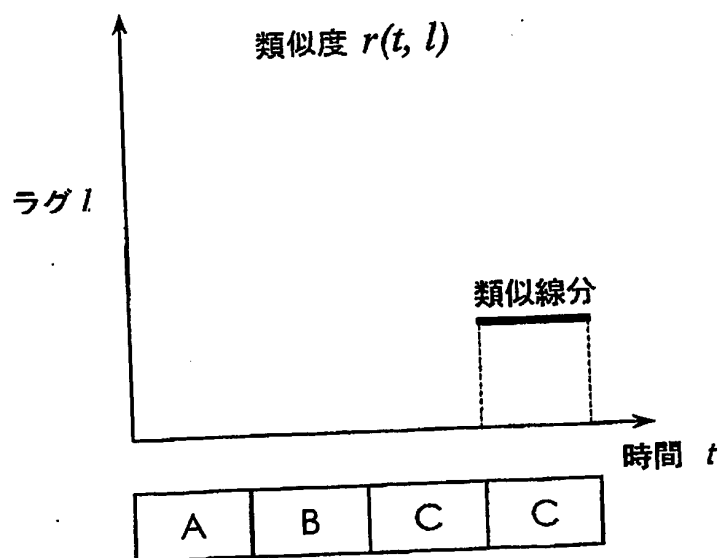




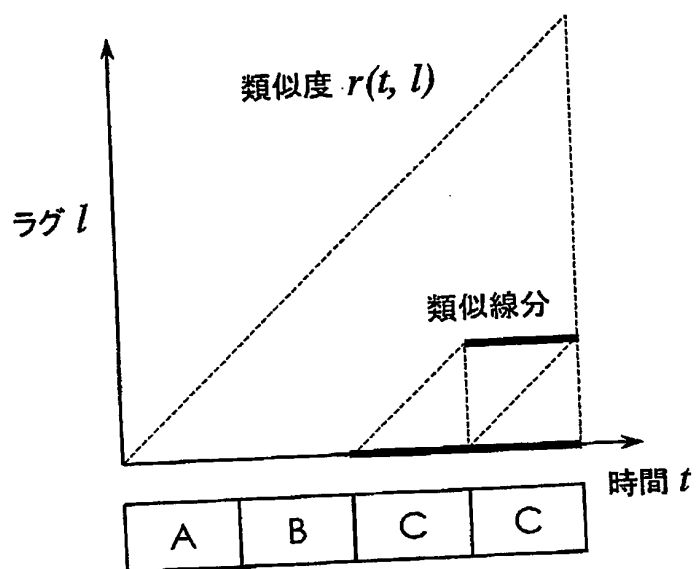
【図 9】



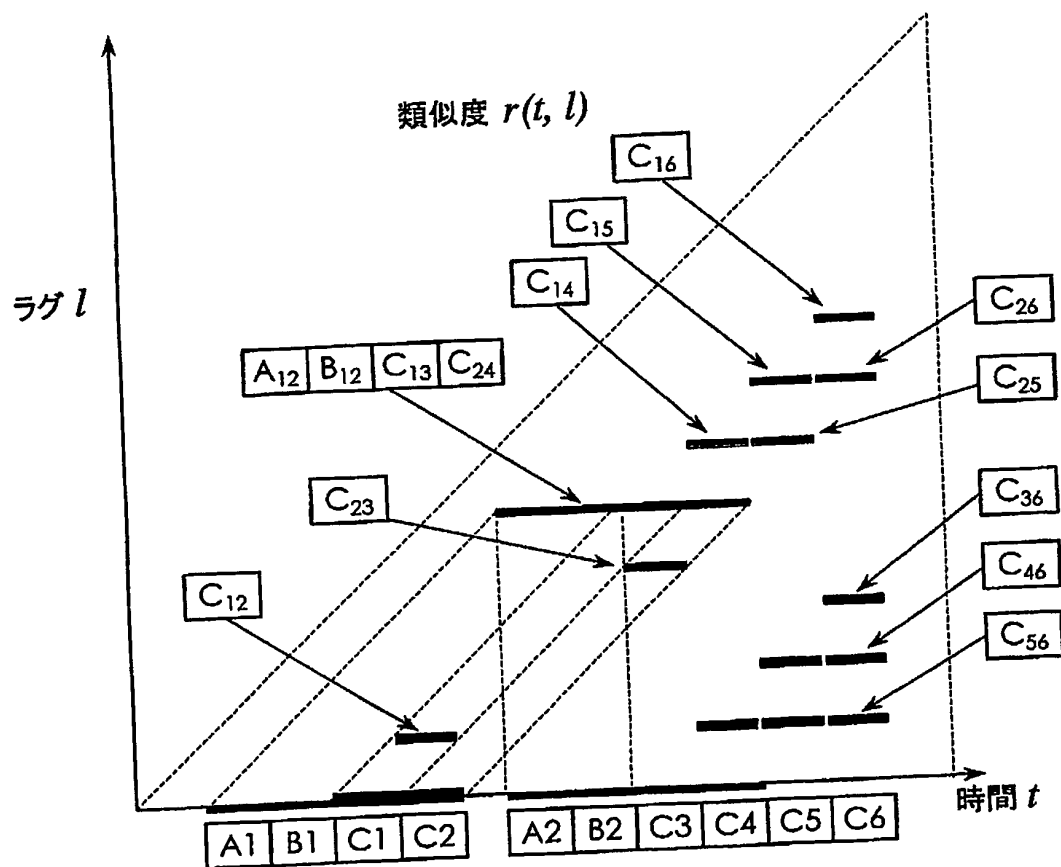
【図 10】



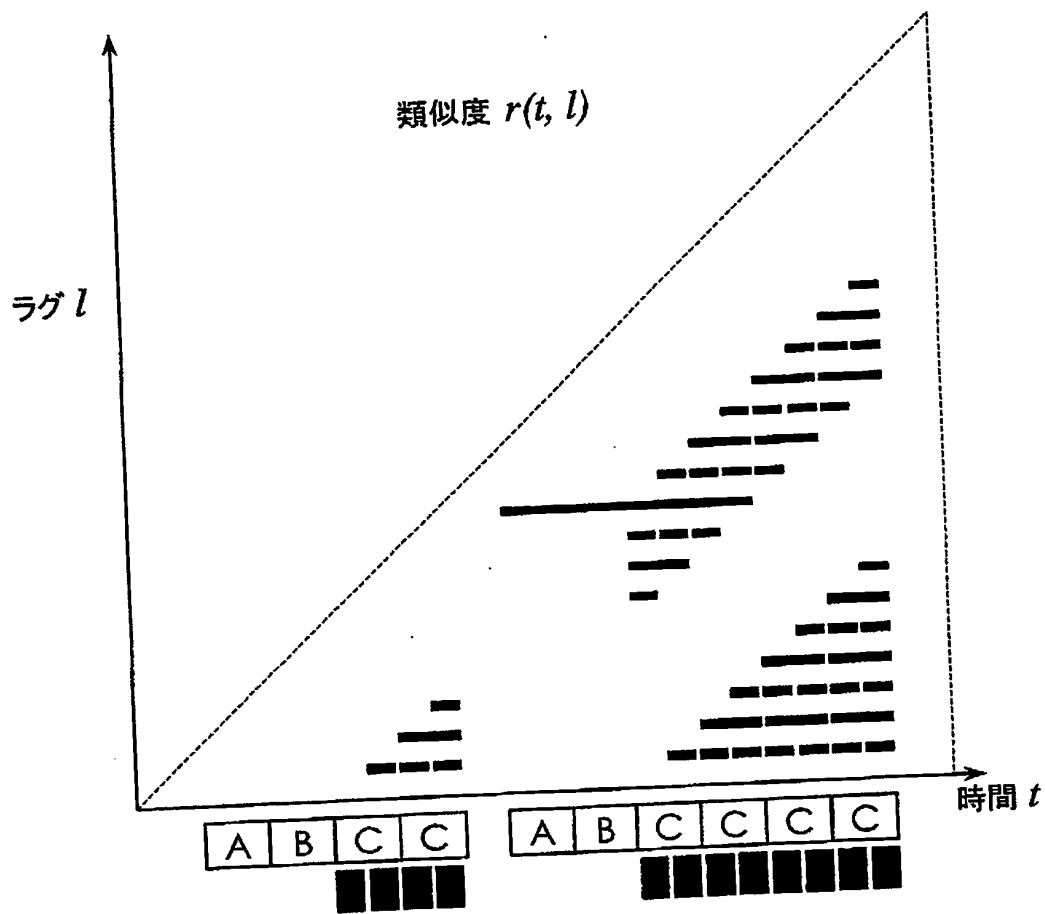
【図 11】



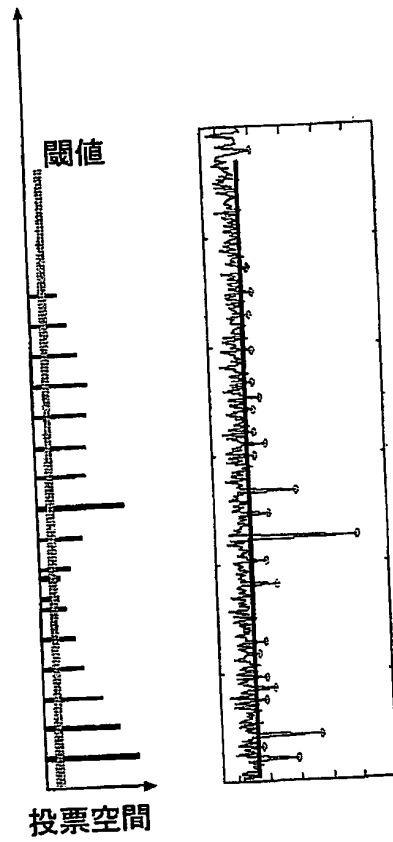
【図 12】



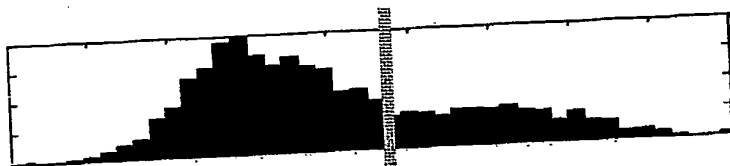
【図 13】



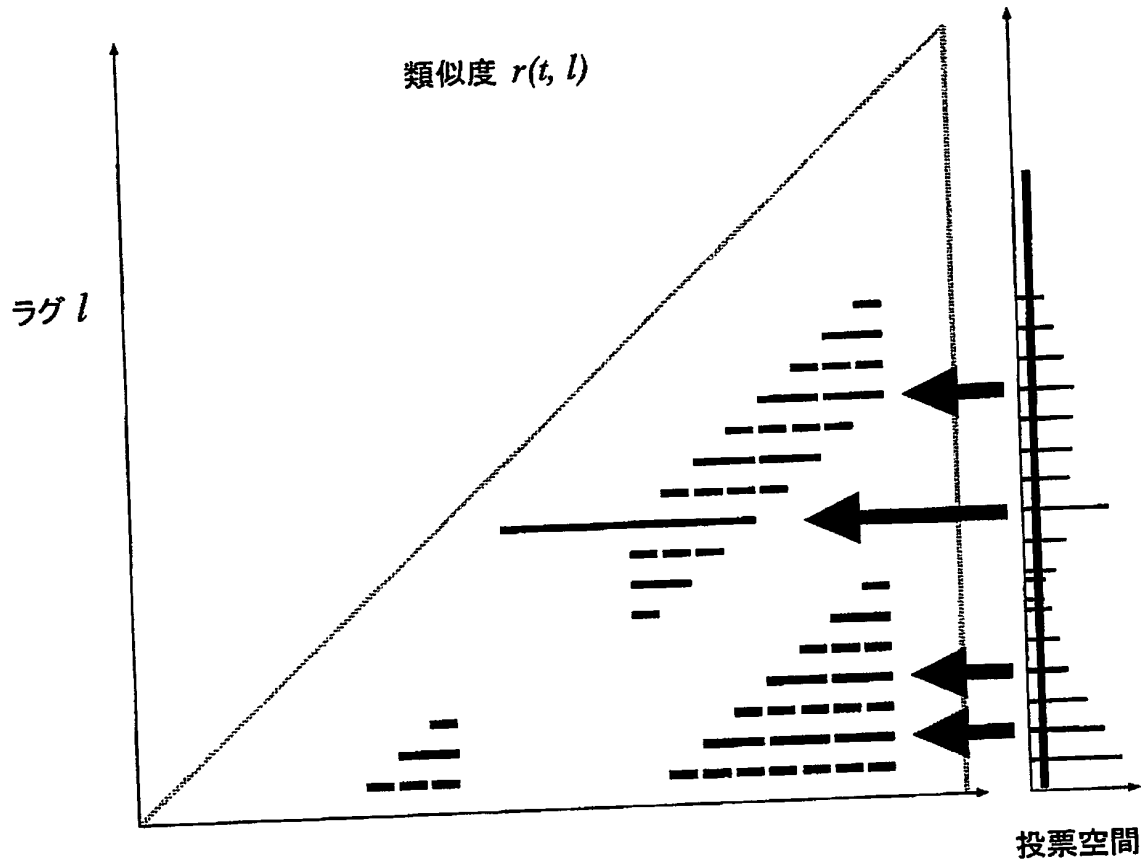
【図 14】



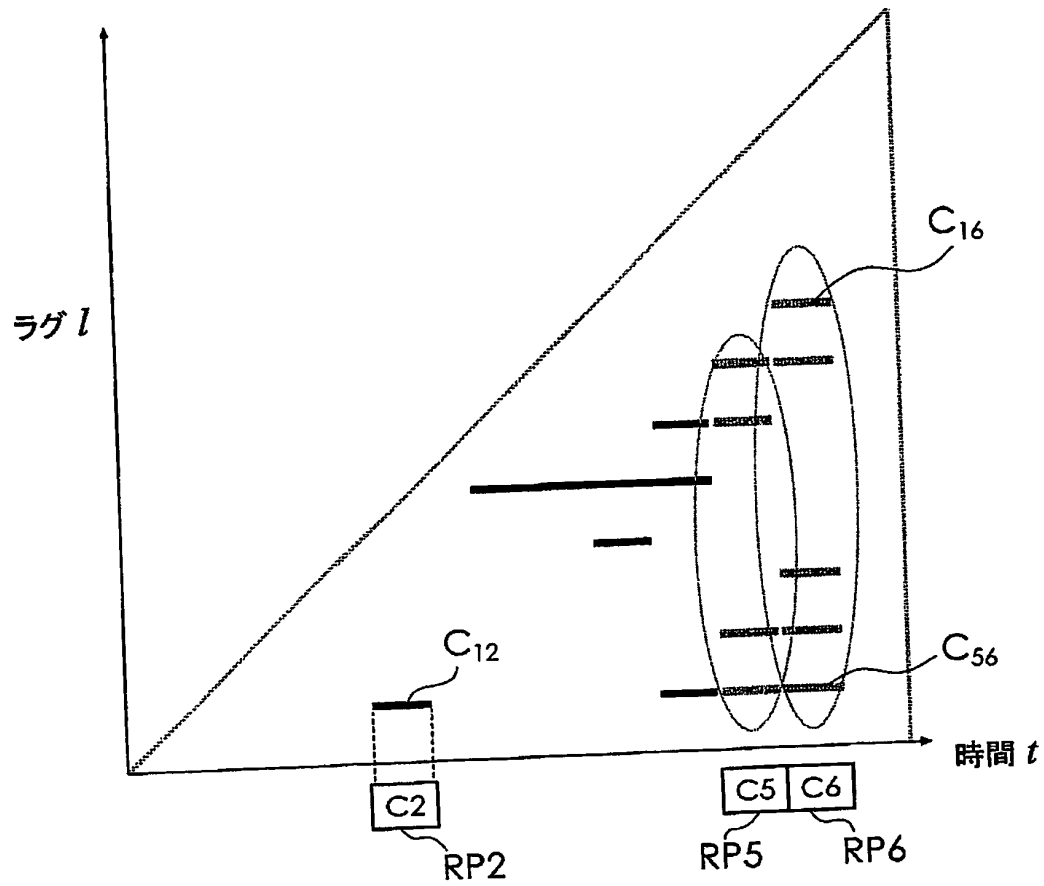
【図 15】



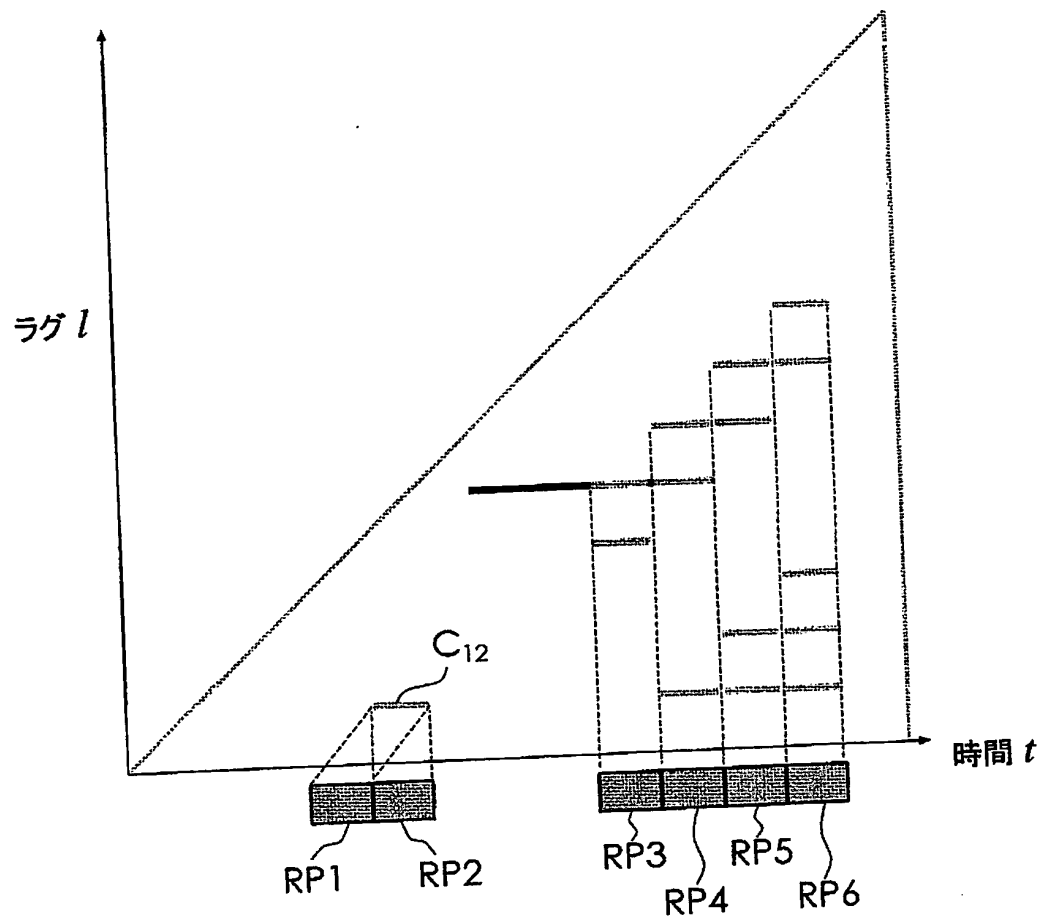
【図 16】



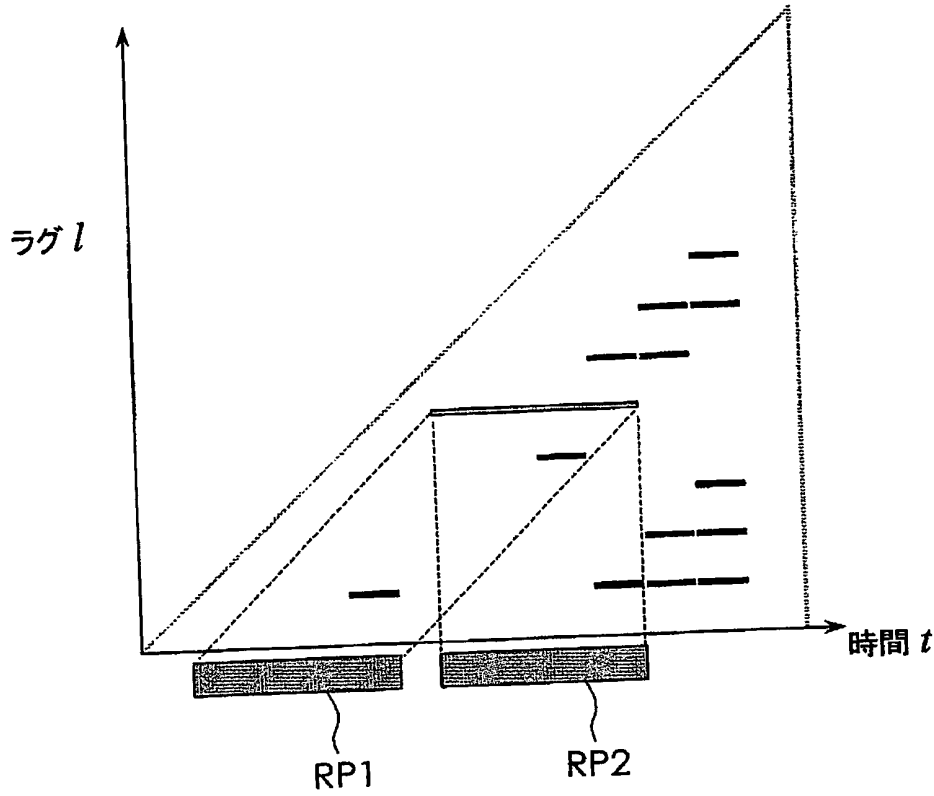
【図 17】



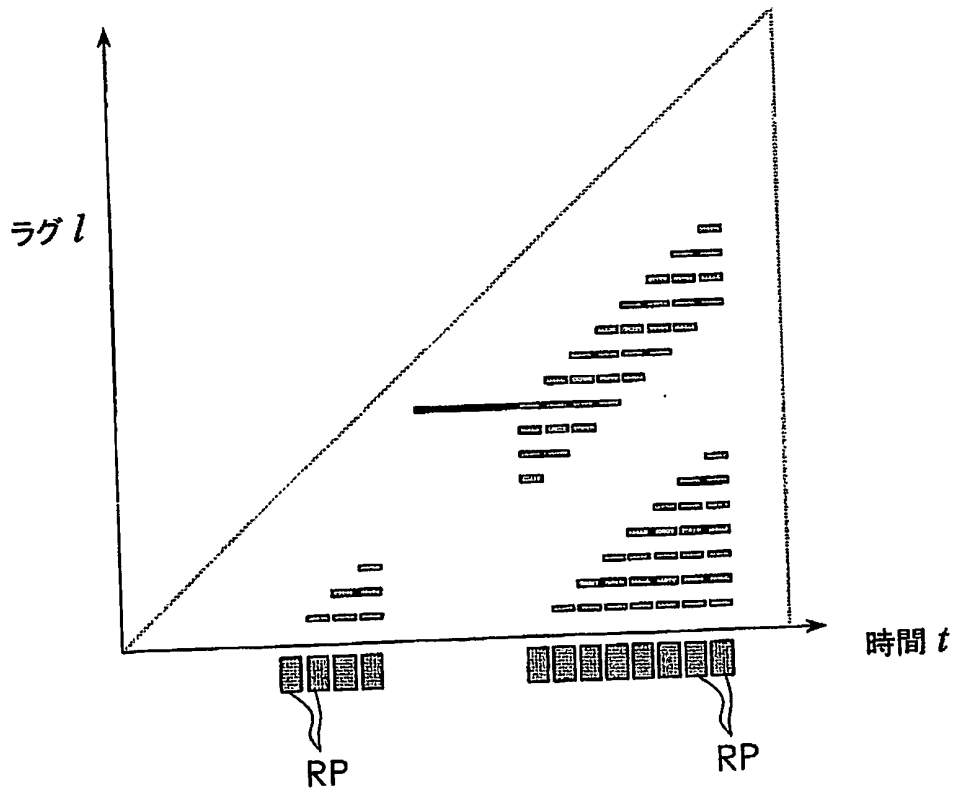
【図 18】



【図 19】

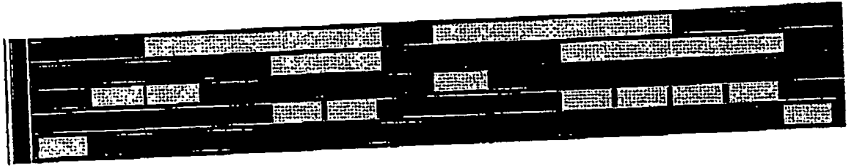


【図 20】

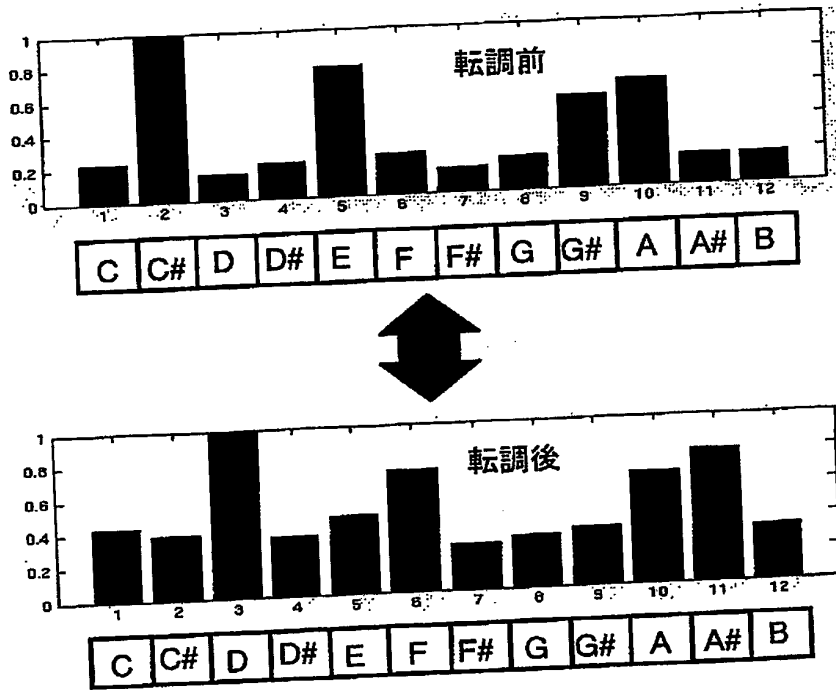




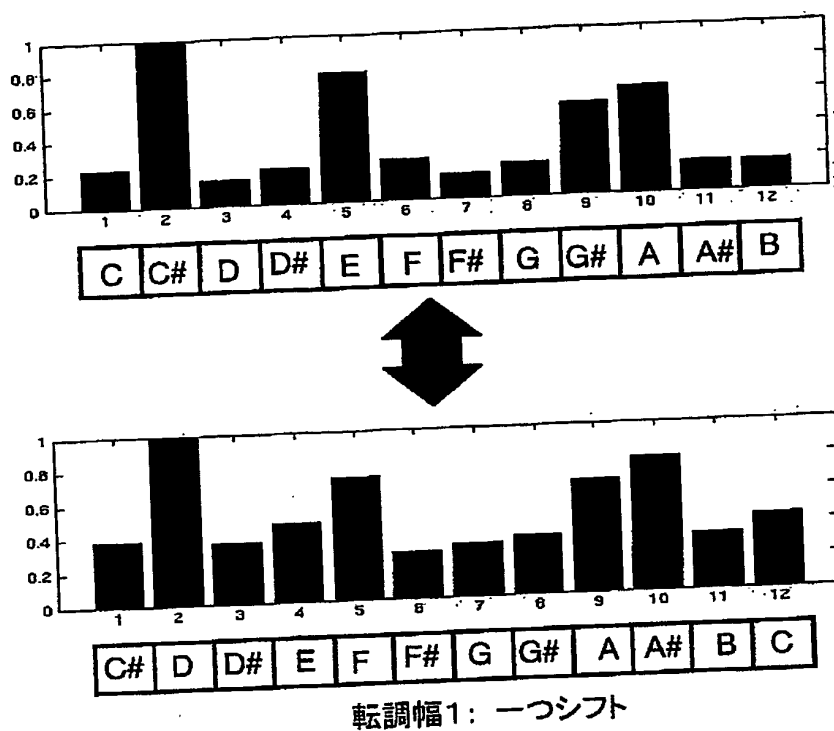
【図 2 1】



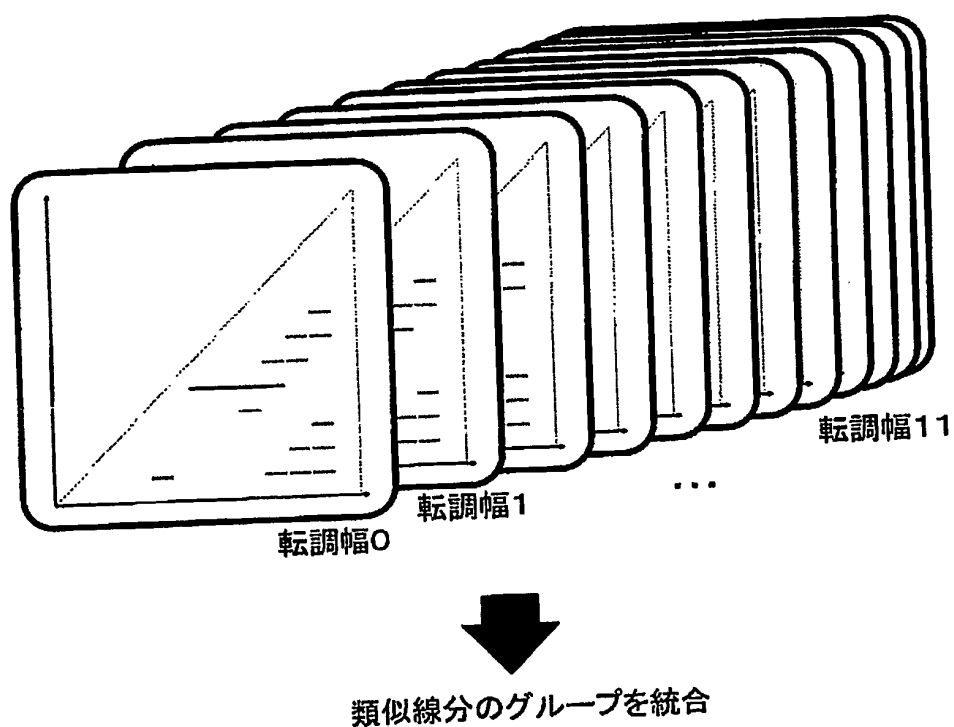
【図 2 2】



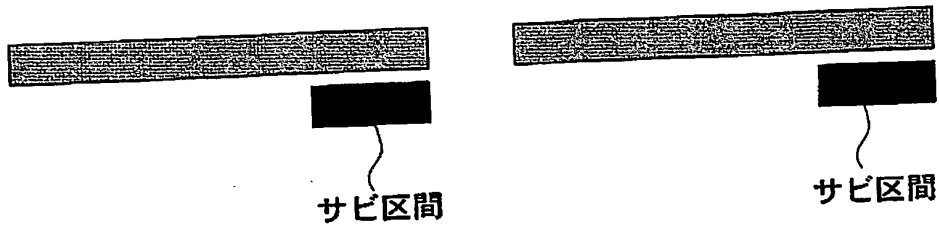
【図 23】



【図 24】



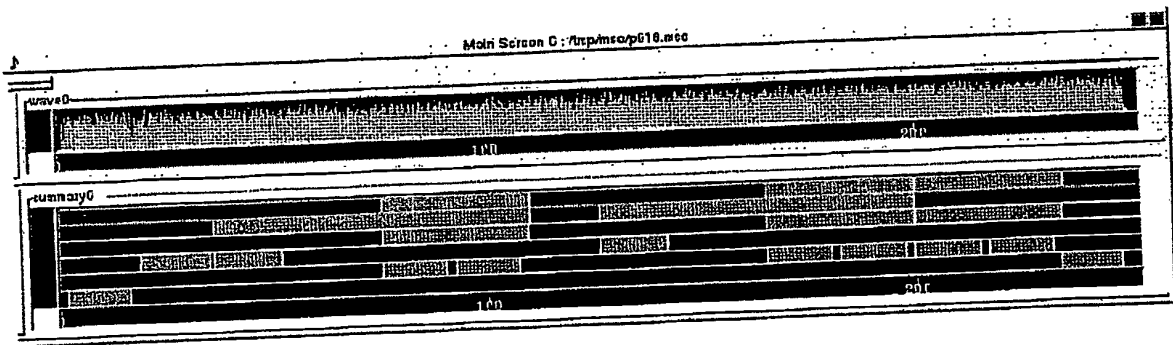
【図 25】



【図 26】



【図 27】



**【書類名】要約書****【要約】**

**【課題】** 楽曲中に出現するすべてのサビの区間を網羅的に検出することができる音楽音響データ中のサビ区間を検出する方法を提供する。

**【解決手段】** 音響信号を入力し、音響特徴量の抽出を行う。音響特徴量間の類似度の計算を行い、繰り返し区間のリストアップを行う。繰り返し区間の統合を行い、転調を伴う繰り返し区間の検出を行って、それらの転調を伴う繰り返し区間の統合を行う。統合した統合繰り返し区間の中から、サビ区間として適切なものを選択する。様々な繰り返し区間の相互関係を調べることで、楽曲中で繰り返されるすべてのサビ区間を網羅的に検出し、それらの開始点と終了点を推定できる。また、転調後でも繰り返しと判断できる類似度を導入することで、転調を伴うサビ区間の検出ができる。

**【選択図】** 図 1

## 認定・付加情報

特許出願の番号	特願 2003-342676
受付番号	50301623724
書類名	特許願
担当官	古田島 千恵子 7288
作成日	平成16年 7月 7日

## &lt;認定情報・付加情報&gt;

## 【特許出願人】

## 【識別番号】

301021533

## 【住所又は居所】

東京都千代田区霞が関1-3-1

## 【氏名又は名称】

独立行政法人産業技術総合研究所

## 【特許出願人】

## 【識別番号】

396020800

## 【住所又は居所】

埼玉県川口市本町4丁目1番8号

## 【氏名又は名称】

科学技術振興事業団

## 【代理人】

申請人

## 【識別番号】

100091443

## 【住所又は居所】

東京都港区虎ノ門1丁目25番5号 虎ノ門34

MTビル9階 西浦特許事務所

## 【氏名又は名称】

西浦 ▲嗣▼晴

【書類名】	出願人名義変更届 (一般承継)
【提出日】	平成15年10月31日
【あて先】	特許庁長官 殿
【事件の表示】	
【出願番号】	特願2003-342676
【承継人】	
【識別番号】	503360115
【住所又は居所】	埼玉県川口市本町四丁目1番8号
【氏名又は名称】	独立行政法人科学技術振興機構
【代表者】	沖村 憲樹
【連絡先】	〒102-8666 東京都千代田区四番町5-3 独立行政法人科学技術振興機構 知的財産戦略室 佐々木吉正 TEL 03-5214-8486 FAX 03-5214-8417
【提出物件の目録】	
【物件名】	権利の承継を証明する書面 1
【援用の表示】	平成15年10月31日付提出の特第許3469156号にかか る一般承継による移転登録申請書に添付のものを援用する。
【物件名】	登記簿謄本 1
【援用の表示】	平成15年10月31日付提出の特第許3469156号にかか る一般承継による移転登録申請書に添付のものを援用する。

特願 2003-342676

出願人履歴情報

識別番号

[301021533]

1. 変更年月日

2001年 4月 2日

[変更理由]

新規登録

住 所

東京都千代田区霞が関1-3-1

氏 名

独立行政法人産業技術総合研究所

特願 2003-342676

出 願 人 履 歴 情 報

識別番号

[396020800]

1. 変更年月日

1998年 2月24日

[変更理由]

名称変更

住 所

埼玉県川口市本町4丁目1番8号

氏 名

科学技術振興事業団



特願 2 0 0 3 - 3 4 2 6 7 6

出 願 人 履 歴 情 報

識別番号

[ 5 0 3 3 6 0 1 1 5 ]

1. 変更年月日  
[変更理由]  
住 所  
氏 名

2 0 0 3 年 1 0 月 1 日  
新規登録  
埼玉県川口市本町4丁目1番8号  
独立行政法人 科学技術振興機構

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**